DOCUMENT RESUME

ED 047 763 LI 002 652

AUTHOR	Kowitz, Gerald T.; And Others
TITLE	General Information Processing System: The GIPSY/ERIC Abstract Retrieval System.
INSTITUTION	Oklahoma Univ., Norman.
PUB DATE	71
NOTE	89p.; Information Science Series Monograph 5
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Educational Resources, *Information Processing, *Information Retrieval, *Information Systems, Search Strategies
IDENTIFIERS	Educational Resources Information Center, ERIC, *GIPSY ERIC Abstract Retrieval System, On Line Systems

ABSTRACT

The General Information Processing System (GIPSY) of the Merrick Computing Center is a user dominated system. It can be used to manipulate and retrieve both numeric and alphabetic material. One of its most interesting uses is in the selection and retrieval of records and documents or of selected entries from records and documents. It has been used for almost two years with the items in the "Research in Education" file, and now also includes the "Current Index to Journals in Education" file. GIPSY does not require extensive or detailed programming, and is capable of responding to the user as he asks a sequence of questions. The report includes: two sample searches, terminal printouts, and terminal displays. (MF)

ED 047763

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

James W. Sweeney

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

GIPSY

GENERAL INFORMATION PROCESSING SYSTEM

The GIPSY/ERIC Abstract Retrieval System

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

by

Gerald T. Kowitz

Ronald R. Reeves

Jerry R. Prather

INFORMATION SCIENCE SERIES

MONOGRAPH NUMBER 5

UNIVERSITY OF OKLAHOMA

NORMAN, OKLAHOMA 73069

I 002 652

© Copyright University of Oklahoma, 1971
Reproduction of this material in whole or in part only by permission.

PREFACE

For the past two years the staff of the Merrick Computing Center and the College of Education of the University of Oklahoma have cooperated in operating a search program of the ERIC records through GIPSY. Many of the searches were conducted for staff and graduate student research; others have been done on a commercial basis for interested agencies. The writing of this document commenced shortly after the first GIPSY/ERIC runs. Its publication has been delayed because of frequent changes in the system. Some of the changes have been a result of additional ERIC records becoming available; others have been the generation of new capabilities of GIPSY.

The operation has used a variety of personnel. Some of the operators are administrative or faculty members of either the Department of Information Sciences, the Merrick Computing Center or the College of Education. Others have been graduate students. A number of secretaries have also become proficient at terminal operation. While the background knowledge and professional skill of the professional operator is stressed in making an efficient GIPSY/ERIC search, we often find that the professional personnel has insufficient typing skills to keep up with his professional knowledge. In such cases it is useful to provide him with a secretary who has a high level of typing skills so that the manipulation of the system can keep pace with his conceptual strides through the records and the system.

The production of this monograph was possible only through the dedicated

assistance of the staff of the Office of Education Special Projects. Mrs. Paulette Bishop was responsible for coordination and production. Miss Emily Ekonen had a major editorial responsibility and Mrs. Linda Plaskett had a primary responsibility for the typing and for the many special figures in the manuscript.

Finally, we must acknowledge the continuing and strong support of Dean Robert E. Ohm of the College of Education and Dr. James W. Sweeney, Director of the Merrick Computing Center, in the operation of the system and the production of this monograph. Mrs. Felicia Reid, Assistant Director of the Merrick Computing Center, provided valuable assistance. Special citations must be given to Mr. Larry Greenwell, Mr. Robert Shields and Mr. Charles Addison for their interest in the potential of the ERIC records when made available through this innovative and fascinating search process.

GTK
JP
RRR
2/2/71

CONTENTS

	Page
PREFACE.	
INTRODUCTION	
Dean Robert E. Ohm, College of Education.	1
Dr. James W. Sweeney, Merrick Computing Center.	3
I. INFORMATION USAGE IN EDUCATION	8
II. THE ERIC SYSTEM.	17
III. THE COMPUTERIZED INFORMATION SYSTEM.	22
IV. BOOLEAN LOGIC: The Basis for Information Manipulation	43
V. SAMPLE SEARCHES.	51
VI. BATCH PROCESSING GIPSY/ERIC.	72
REFERENCES	74
APPENDIX	
A Current ERIC Clearinghouses.	76
B How to Request a GIPSY/ERIC Search	80

FIGURES

No.	Title	Page
	Flowchart of Retrieval Dialogue.	7
1.	An ERIC Record	20
2.	The "Help" Command	24
3.	The "Define" Command	26
4.	An ERIC Record	27
5.	The "???" Command.	28
6.	Flowchart of the GIPSY/ERIC Search Process	42
7.	Initial Questions for a GIPSY/ERIC Search.	52
8.	First Sequence of a Simple Search.	53
9.	Second Sequence of a Simple Search	56
10.	Third Sequence of a Simple Search.	58
11.	Fourth Sequence of a Simple Search	59
12.	Final Sequence in a Simple Search.	61
13.	First Sequence of a Complex Search	63
14.	Second Sequence of a Complex Search.	66
15.	Third Sequence of a Complex Search	67
16.	Printout at the Terminal	68
17.	Display at the Terminal.	70
18.	Final Sequence in a Complex Search	71

DIAGRAMS

No.	Title	Page
I.	Logic A.	45
II.	Logic A or B	45
III.	Logic A and B.	45
IV.	Logic A and not B.	47
V.	Logic not A.	47

INTRODUCTION

Comments by
Dean Robert E. Ohm
College of Education

The advent of large, alpha-numeric data bases designed for computer retrieval and stored on disc or tape is a recent phenomenon in management and education. Up to this point the computer as a management and research tool was almost entirely given over to "number crunching" with emphasis on the development of new and more sophisticated programming languages to treat and manipulate numerical data. However, new data bases that are repositories of words, sentences, paragraphs and concepts are becoming available; in short, abstracts of reports, research, articles and monographs are now computer retrievable. We are now at a point of engaging in massive "word crunching" comparable to computer capability with numbers. The prospects for the soft sciences now begin to approach those of the number-based sciences.

The breakthrough was accomplished by the invention of a new language, the General Information Processing System, that enables users to engage in dynamic interaction with a large data base through ordinary English without the need for intermediate programming.

Using a computer-based retrieval system, a researcher can now interact with a much larger data base than was previously possible. For example, the user can now search the 32,262 abstracts of the current ERIC base and retrieve what he needs within an hour. User problems in coping with the immediacy of full information on a problem or topic have

yet to be investigated. The consequences of this for research and publication are not predictable. As other "soft science" data bases such as the School Administration Abstracts, Child Development Abstracts or Psychological Abstracts become available on tape, new skills in information use will need to be developed and new user training programs will need to be constructed. New information theories will need to be developed to guide research and training activities with particular emphases on feedback for the improvement of the abstracting processes so that the abstracts contain the high information density required for research and application purposes.

Similarly, management decision making will face a new challenge as the manager interacts directly and immediately with a data base. New patterns of training in the use of information, in problem solving and in decision making will need to be developed to enable the manager to use this new form of access effectively. The frequently asserted but only partially demonstrated notion of a total information system for an organization is now within the realm of possibility. It will be possible to test the theory of organization as an information system and the practical boundaries of the term "total." In summary, we look forward to some years of exciting theoretical and practical research and development activity in information science.

Comments by
Dr. James W. Sweeney, Director
Merrick Computing Center

GIPSY (General Information Processing SYstem) is a question-oriented General Information Storage and Retrieval System developed at the University of Oklahoma Computing Center. It was designed as a flexible, user-oriented system, for the collection, maintenance and retrieval of information and has found applications in a wide variety of disciplines.

GIPSY evolved from the realization that most traditional methods of information retrieval are limited in their scope, and complex in their operation. Therefore, GIPSY was designed with facilities to allow a full spectrum of information to be stored and easily retrieved by the user. This is accomplished through a highly flexible, easily learned, retrieval language which provides direct, interactive communication between the user and his data base.

As GIPSY does not establish preconceived relationships between the information collected, it permits the user to answer "ad hoc" inquiries and also to "browse" through his information collection. This is accomplished by making the entire content of each document available as selection criteria and not merely some predetermined keywords or index terms. This capability not only provides unlimited access to the data base, but also saves the costly and complex rigors of reprogramming to answer particular questions. In general, the system makes possible exhaustive inquiries which can be accomplished in minutes rather than weeks or months.

Design Objectives

The design objectives of GIPSY can be succinctly described by the concepts of flexibility, adaptability and simplicity. The flexibility of the system is apparent in its ability to store, maintain and retrieve variable size records composed of a variable number of items. This allows the user to assemble diverse types of information and thus adapt the system to his particular needs. In addition, the variable nature of the record structure provides for more economical utilization of storage facilities.

The flexibility of the system also permits adaptation to a number of problems. The system has found applications in a wide variety of disciplines including government, research and education. As GIPSY may be used in either a batch or teleprocessing mode, the system is compatible with most IBM 360 configurations meeting the minimum requirement of 82K bytes of storage, the decimal arithmetic feature, full operating system (OS), and two of the 2311's and 2314's, or equivalent direct access space. The teleprocessing version supports either the IBM 2741, the Datel 30 or equivalent units.

No matter how sophisticated a retrieval system, its utility will depend on whether it can be utilized by the layman. GIPSY was designed to make the communication between man and machine function in a clear, concise, English-like language which is easily mastered by the non computer-oriented person. Thus, GIPSY presents a method of communication in a form that does not suffer from the ambiguities present in plain English, yet does not require an intermediate programmer between the user and his information.

The Retrieval Language

QUESTRAN (Question Translator), the retrieval language of GIPSY, is a non-procedural language implemented with the user in mind. The language is of macro type, consisting of commands to control the program modules, and parameters which initiate specific operations. It was designed to have as few syntactical restrictions as possible in order to retain a clear, logical and concise structure. Experience has proven that a non computer-oriented person, familiar with a particular discipline, can begin to get useful answers within 15-30 minutes using the simple command structure of QUESTRAN.

QUESTRAN provides the ability to answer "ad hoc" inquiries and also allows the user to extend the normal scientific method to computerized information retrieval in much the same manner that research is accomplished in the library. A hypothesis is expressed in the form of a question posed to the data base. The user identifies the variables or conditions he is seeking and the logical relationship to be satisfied (SELECT). After the search has been accomplished, the results are displayed so that the questioner may analyze them. The selected records can then be submitted to one of the processors or the user may refine this subset through further inquiry (ITERATE). This ability was implemented in the SELECT/ITERATE command of the retrieval language and is illustrated in the following flowchart. Utilizing a selected record file (SRF), the user has an option after each SELECT/ITERATE to either process the selected records, or place further constraints on the selected subset, thereby attempting to retrieve a subset which is closer to his desired results. The subsets created as a result of this process

are numbered for identification purposes. This provides the user with the ability to return to any previously created subset to redefine or initiate new constraints.

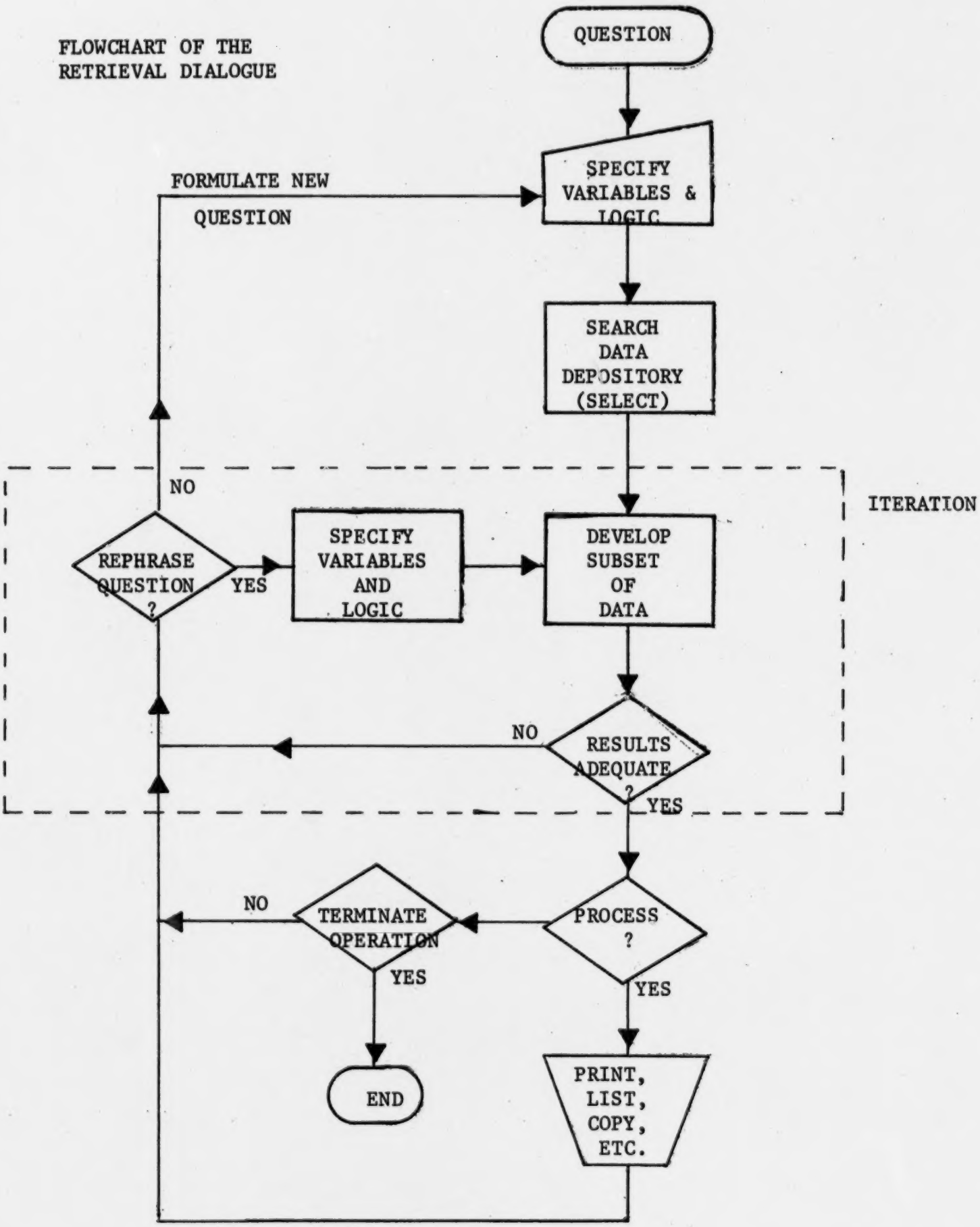
The system provides the user with the ability to utilize any part of the record in the selection criteria and to use full Boolean logic (AND, OR, NOT) to describe very concisely the relationship the records must satisfy for selection. These logical relationships can vary from the very simple to the very complex, and are therefore readily adaptable to the skills of the researcher. This flexibility and power, combined with its simple method of communication, make GIPSY a highly useful retrieval mechanism.

Processors

GIPSY provides the user with a number of processors which he may apply to the records selected after an inquiry. These processors include a printing capability which allows all or part of each record to be printed (PRINT, LIST); a method for obtaining summations and averages of given items (SUM); and a sorting capability (SORT). In addition, the user is provided with a processor that enables him to create standard, fixed field, fixed length records from GIPSY free form records, based on parameters established by the user (COPY). This processor provides interface between GIPSY and established fixed field systems.

The current processors are adaptable to a variety of uses and provides a flexible, adaptable package for information handling. As the system is modular in design, additional processors can be developed and installed to meet specific needs without major reprogramming.

FLOWCHART OF THE
RETRIEVAL DIALOGUE



PART I

INFORMATION USAGE IN EDUCATION

Education, like other professional activities and large organizations in our technological society, depends heavily upon information. However, in education information has a unique role, since from one point of view every academic area has a vested interest in the processes of schooling. At some point its content must be taught and new workers prepared for the field. Information as a part of the educative process can be examined under four general headings.

1. Planning
2. Management
3. Development
4. Research

Each of these has much in common with similar activities in other fields, but each also has unique aspects.

Planning

Planning new programs is a major executive activity. New programs are required because of changes in the society which supports the school and because of advances within the profession. The development of technology in the last half century has placed many new demands upon the school system. It has also altered some of the traditional goals of the school. The general trend has been to require more education and training and a higher level of quality in instructional programs. The growing affluence

of our society has had a different impact. On the one hand, it has brought increasing concern over the nature of the student, his personal and social development. Most professionals perceive these as additional or supplemental goals, not as substitutes for the traditional academic goals of the school.

The increasing affluence of the modern world has emphasized a reexamination of many traditional goals and procedures. Many school processes which were perceived as necessary expedients because of the financial problems in school support can now be reexamined. Budget regulated planning is no longer regarded as a fundamental rule in school administration, as it was only a few years ago. Today many educational problems which have persisted over the years can be examined on the professional level rather than disposed of on the administrative level because of budgetary constraints.

The changes require much planning on the part of school executives. Information has a major role in planning. On the one hand, it is a source of new ideas; and on the other, it avoids the wasteful duplication of developing programs and procedures that have already been well established in other areas.

Ideas for new programs, and the planning needed to make them operational, do not spring into being from nowhere. They require professional knowledge and experience.

Management

Traditionally, management of school programs has focused upon the identification of philosophical goals for the operation. While this work is still important, modern management tends to work more closely with

reality. It is concerned with identification of objectives which can be specified in operational or behavioral terms. From another point of view, it is concerned with identifying alternate pathways for achieving the goals. The analysis which identifies feasible alternatives is, indeed, a major process in educational management.

The second stage in management concerns the development of an iterative loop. Through the loop, information is returned to the manager from the operation. He can use the information in evaluating the effectiveness of his plans. By comparing the product with the anticipated results, he can arrive at a judgment about the effectiveness of his operation. More important, he can move forward to make adjustments, some minor and some major, in the operational process to insure a better agreement between the goals and the products. The processes of receiving information, of evaluating and of replanning are the keys to the improvement of school operations.

Development

Planning and management are continuing activities. The development of new programs or new processes is based upon them. Development may be perceived as a second stage of the planning and management activities. It relies heavily upon what others have learned and upon a feedback of information on how the present program is working. It is unique in that it includes two new processes: First, it is sensitive to emerging alternatives, to possibilities which were not available when the initial planning was done. Second, it is sensitive to unanticipated advantages and to new achievements from the operation. These are often referred to as "spin-offs" of the operation. In some instances, the unanticipated products may

achieve importance equal to or even greater than that of the basic program.

Finally, development of a program encompasses the processes of demonstration and dissemination of new concepts. Alert school administrators are always searching for new ideas and better programs; they are seeking information. As soon as a program begins to produce tangible results, the local school will begin to receive requests for information about it. Often the request will ask permission to send a visiting team to view the program. Two problems emerge immediately:

The presence of visitors, especially in large numbers can interfere with the operation of the new program--and even with the entire school.

The local leaders must provide the visitors with information about what they should look for. If they do not, the visitors may not perceive the activities, processes and facilities that make the difference. Instead, they will see only superficial resemblances to other programs which they have observed. Information must be provided with care and planning or the observers will come to believe that they are being provided with a show and denied an opportunity to observe the program as it really operates.

Because a program or process is successful in one location does not mean that it can be transferred and installed as an identical operation somewhere else. Demonstration in new locations always requires minor adjustments. These adjustments also presume a feedback of information to the program manager. He must know what impact the adjustments have upon the operation and the product.

Dissemination is a further refinement of the process of educational development and improvement. Until information about a change is made

available to other workers in the field, it lacks a potential for making a major contribution to the operation of schools. Because of minor variations resulting from differences in local schools, dissemination procedures must include information about adjustments that have been made, those that can be made, and those that should not be made. Thus, the dissemination aspect of development depends heavily upon the processes of development and management. Unless sufficient information is generated about the basic nature of the program, local operators may make changes which introduce inefficiency or which actually reduce the potential of the new operation.

Research

Research is an honored tradition in the modern technological world. It focuses upon the generation of new information. It may or may not have any immediate, practical application. The traditional division between pure and applied research is no longer as clear as it once was. However, there must be effective channels of information between those concerned with finding solutions to the immediate operational problems and those concerned with resolving inconsistencies among beliefs and practices in the field of education.

The flow of information is fundamental to the operation of a research project. The project begins with an attempt to specify the problem upon which it will focus. Information about similar or related projects can help specify the problems. Reports on other projects can also be used to identify techniques, instruments and evaluative designs for the operation. A broad base of data can provide the researcher with many ideas and specific routines or instruments to build a model of the operation which he wishes to study. The process of model building is essential to research;

the final product of the research can be no better than the model upon which it was based. Finally, the information from the field about other attempts will reveal to the researcher incompatibilities of ideas and contradictions in techniques. These are prime sources of future research projects.

The unique role of information in education can be seen along another dimension.

Education as a Flow of Information

The processes of schooling, functioning as they do with the maturing youth, complicate the use of information in education. Few fields dare wait until the student is mature to introduce him to their speciality. It is necessary, therefore, to filter some selected parts of the academic fields, to restyle them into school subjects. The style must be such that it does not form barriers to the child's progress; hopefully, it will also develop his interests in the several fields. Thus, every academic field must consider the problems of pre-college education and make some contribution to its design and operation. One unusual result is that information relevant to education must be regarded as infinite. As such, problems of information processing, of storage, manipulation, retrieval and transfer are major activities in education.

During the 1950's and the early 1960's the major movement in the public schools was the revision of standard curriculum content. Specialists in education and specialists in academic fields were urged time and again to work together more closely in order to develop an improved curriculum for the public schools (Bruner, 1960). One result was an exponential expansion of the professional literature.

From a second point of view, education draws heavily upon the professional literature of the behavioral and social sciences. This is especially true where the focus is upon the design and operation of innovative programs. Strangely, while much of the content of these sciences have points of obvious relevance, they may still have little to say about the operating procedures of the school. It is important to distinguish between the processes of education and the processes of schooling (Stephens, 1967). A common error, in fact, is the failure to recognize the constraints inherent in the processes of schooling, many of which are largely irrelevant to the process of education. The history of school operation is much longer than the history of any topic that is taught in the schools today. Many customs and traditions of school operation tend to continue independent of changes in the content and method of teaching.

A third point of contact between information and education arises from recent attempts to introduce principles of modern management into school operation. The traditional patterns of management and operation were based upon a concept of a small, independent school. State education departments and other government units which had direct contact with the schools often recruited their staff from small districts. One result was a tendency to perpetuate the traditional patterns of operating small schools. Urbanization, and the resulting growth in the size of school districts, forced a recognition of the growing dependence of the schools on government operations and interrelationships with community agencies. Schools are no longer small, independent organizations which can operate on an informal basis. Traditional patterns for operation are being questioned. Modern management techniques are dependent upon the free and

efficient flow of information. Blai (1970) pointed out that the key to modern management techniques in education is the efficient use of information in educational decisions. The information upon which decisions are to be made must be relevant and timely within the decision setting.

Not too long ago, a local school administrator would automatically take pride in instituting a new instructional program; his main evaluative questions would concern its philosophical consistency and its financial feasibility. Today he is asked to examine the results of the programs in terms of educational gains as well as in terms of fiscal efficiency, philosophy and politics.

Traditionally, the administrator could rationalize some duplication of effort by insisting that his school district was significantly different and, therefore, even though his program superficially resembled that of another district, it operated under different conditions and had different requirements. With the publication of many documents giving descriptions of programs, it has become increasingly difficult for the administrator to defend the idea that his operation is significantly different. Also with more comprehensive reporting came the view that some inefficiency was inherent in the development of new programs; any advantage that could be gained from reviewing other programs was important; the ability to manage information gained new significance in education.

Organizing Information

Although the history of educational innovation is a long one, at least as long as the history of the schools, the process of change and improvement in school practices gained publicity beginning in late 1957. The launching of the Russian Sputnik in October, 1957, caused many people

to ask why the first satellite was Russian, not American. A popular answer was that our schools had failed to produce a large enough quantity of sufficiently competent scientists and mathematicians. One response was a series of laws, beginning with the National Defense Education Act and continuing through the Education Professions Development Act, which provided funds to encourage and support the development of new techniques in education. Since that time, the number of documents with some significance to professional education increased astronomically. It soon became obvious that a central effort would be needed to provide for the management of information in education. An early attempt was the Educational Resources Information Center (ERIC), which is probably the largest single information network in the world. Its operation has been a definite asset to education and has called attention to many previously unrecognized problems of information in education.

The problem of information retrieval is clearly a function of the size of the data base. As the base gets larger, the probability of being unable to locate a specific item increases. The inability to retrieve information quickly and efficiently can frustrate the user. A common reaction is for him to condemn the data base, when, in fact, it is his approach to the information which is inadequate. Obviously, information cannot be used until it has been retrieved. In fact, the retrieval of information is only the first problem which the user must face. However, unless it is resolved satisfactorily, all subsequent applications will be impaired.

PART II

THE ERIC SYSTEM

The Educational Resources Information Center (ERIC) is a network of information activities supported by the U. S. Office of Education (USOE). ERIC provides three approaches to information: Research in Education (RIE), the ERIC Document Reproduction Service (EDRS), and Current Index to Journals in Education (CIJE).

ERIC was designed to provide organization and management and, thereby, access to a great number of documents relevant to education. Hopefully, it would minimize the duplication of effort, disseminate useful approaches for wider applications and stimulate thinking about better methods and procedures. Information is thus perceived as a useful end in itself, as an active ingredient in educational planning and management, and as a catalyst for innovation.

ERIC

The ERIC network consists of 20 clearinghouses, each of which is assigned to collect and screen documents on a particular topic in education. Most of the clearinghouses are located at universities, although some are operated by professional organizations and special interest groups. Appendix A presents a list of current clearinghouses.

The more significant publications are abstracted, and both the abstract and the document are forwarded to ERIC Central; the remainder of the documents are retained at the clearinghouse and are available from that source.

The clearinghouses operate under contracts with USOE. From time to time new clearinghouses may be established or existing ones modified or moved to a new location. The flexibility is designed to provide current coverage of topics of concern in education.

The ERIC network also contracts with a commercial firm, National Cash Register Company, to provide microfiche and hard copy of the complete document. The reproduction service (EDRS) is a commercial operation. In some cases, the entry in RIE will indicate that the document is not available through EDRS. When this occurs, there will usually be a different source listed where the document can be obtained.

ERIC has a contract with another commercial firm, LEASCO Systems and Research Corporation, to provide computer tape containing the RIE and CIJE entries.

RIE

The abstracts forwarded to ERIC Central are entered into RIE, a printed index system, and are also placed on computer tape. A user of the system may search for useful documents through RIE, a paperback volume which is issued each month. Semiannual and annual indexes are also available. RIE employs a classical indexing system with cross-references by institution, author, and subject. The user also may search the abstracts through a computerized retrieval system. When he has located those references in which he is interested, he can, by use of the ED number, locate the complete document on microfiche or as printed hard copy. The printed hard copy may be ordered from EDRS following the instructions given in RIE.

Since the contents of RIE are in the public domain, only a few copyrighted research reports or articles from professional journals are included in it. However, reports on nearly all of the research projects, demonstrations, programs and conferences sponsored by the U. S. Office of Education are in the ERIC system.

CIJE

In an effort to provide coverage of the mainstream of professional literature, CIJE was initiated. It presents the titles, author and bibliographical reference to journals which have agreed to participate in the program. Over 530 professional publications are in the index. CIJE also provides a brief annotation describing the contents of some of the articles. Each article is classified by descriptors based upon the RIE index system.

Retrieval Processes

As the number of documents in an information system increases, the problems of retrieval increase exponentially. The base of the problem is that each document must be indexed and cross-indexed before it can be retrieved efficiently. The need for many cross-references quickly causes the index system to exceed the main body of data in bulk. The ERIC system, including both RIE and CIJE, for example, is increasing at a rate of around 1,200 documents per month. As a result, the system of indexing is growing rapidly and the efficiency of a classical indexing system is decreasing rapidly. Figure 1 presents, as an exhibit, one ERIC record in which the index system is almost as large as the entire abstract.

Figure 1

An ERIC Record

ACCESSION NUMBER - ED011929

CLEARINGHOUSE ACCESSION NUMBER - VT001550
PUBLICATION DATE -66
INSTITUTION CODE - CIQ11100
EDRS PRICE - EDRS PRICE MF-$0.09 HC-$1.88
DESCRIPTIVE NOTE - 47P.
ISSUE - DH6667
INSTITUTION NAME - CALIFORNIA STATE DEPT. OF EDUCATION, SACRAMENTO.

TITLE - VOCATIONAL EDUCATION FOR PERSONS WITH SPECIAL NEEDS.

ABSTRACT - A SURVEY OF RESEARCH WAS MADE TO IDENTIFY THE VOCATIONAL NEEDS OF INDIVIDUALS WITH SPECIAL PROBLEMS. COMPLETED STUDIES, RESEARCH-IN-PROGRESS, AND CURRENT PROPOSALS WERE INCLUDED. TOPICS COVERED IN THIS SUMMARY ARE (1) AGING, (2) CONTINUATION EDUCATION, (3) THE CULTURALLY DEPRIVED, (4) DELINQUENCY, (5) DROPOUT, (6) EMOTIONAL DISTURBANCE, (7) LOW ABILITY, (8) MENTAL RETARDATION, (9) MINORITY RACIAL GROUPS, (10) PHYSICALLY HANDICAPPED STUDENTS, (11) SMALL SCHOOLS, (12) SOCIOECONOMIC STUDIES, (13) UNDERACHIEVERS, AND (14) WORK OPPORTUNITIES. A BIBLIOGRAPHY IS INCLUDED. (MS)

DESCRIPTOR - ABILITY: ADULT EDUCATION: BIBLIOGRAPHIES: CULTURALLY DISADVANTAGED: DELINQUENTS: DROPOUT IDENTIFICATION: DROPOUT PREVENTION: DROPOUT RESEARCH: *EDUCATIONAL NEEDS: *EDUCATIONAL PROBLEMS: *EDUCATIONAL RESEARCH: EMOTIONALLY DISTURBED: EMPLOYMENT OPPORTUNITIES: MENTALLY HANDICAPPED: MIDDLE AGED: MINORITY GROUPS: OLDER ADULTS: PHYSICALLY HANDICAPPED: POTENTIAL DROPOUTS: SMALL SCHOOLS: SOCIOECONOMIC STATUS: *SPECIAL EDUCATION: UNDERACHIEVERS: *VOCATIONAL EDUCATION

Several attempts have been made to devise new and more efficient systems for building an index. The KeyWord-In-Context (KWIC) index developed by IBM is one example. The use of the inverted and rotated files for indexing permits greater accuracy in locating specific items. However, it also increases the bulk of the index system.

A number of efforts have been made to apply high-speed electronic computers to the problems of search and retrieval of information from abstracting systems such as that provided by ERIC. This monograph describes one such system.

PART III

THE COMPUTERIZED INFORMATION SYSTEM

The General Information Processing SYstem (GIPSY) of the Merrick Computing Center is a user dominated system. It can be used to manipulate and retrieve both numeric and alphabetic material. One of its most interesting uses is in the selection and retrieval of records and documents or of selected entries from records and documents. It has been used for almost two years with the items in the RIE file. It now contains both the RIE and CIJE materials.

GIPSY does not require extensive or detailed programming. Its commands and its outputs are in English and can be mastered in a matter of minutes by most professional users. The most significant function of GIPSY is that it is capable of responding to the user as he asks a sequence of questions. Thus, the professional training of the user and his experience in the field are far more important to a successful search than his knowledge of computers or information systems. While his knowledge of the professional field is his most significant source of assistance, a knowledge of characteristics of the data base is also helpful. This, however, can be acquired quickly by working with the system. Because of its unique characteristics, GIPSY permits the user to engage in reflective or introspective research.

GIPSY does not establish a fixed relationship between the information collected and the user through an index system. It provides him with access to the entire content of each document; he is free to browse and to pursue ad hoc inquiries.

The following pages of this monograph describe the operation of the GIPSY/ERIC search system from a remote terminal. At present, a Datel 30 remote terminal is being used in the College of Education Building through a standard office telephone. Batch processing is also possible and will be considered in a later section.

Self-instructing Commands

The system will respond to a set of commands which can be used in manipulation of the records. Through these commands, the system is, to some extent, a self-instructing system. Three commands may be used to learn about the system: "Help," "Define" and "???." In addition, a Coursewriter III program exists that will teach the operation of GIPSY and the specific GIPSY commands.

"HELP." When the code word "help" is typed, it will cause a list of all commands through which the system is operated to be printed at the terminal. Each command is accompanied by a brief description or definition of its function. A copy of the procedure for using it and the response to the "help" command is presented in Figure 2.

In passing, we should note that in the interaction between the user at the terminal and the machine, the inputs from the terminal are always in lower case letters; the outputs from the computer are in upper case. Where only a signal, rather than information, is required in response to the user's message, a sequence of seven dashes is printed out. This response may be seen immediately below the command "help" in Figure 2.

"DEFINE." The command "define" will print at the terminal the dictionary of abbreviations which can be used with a pre-identified data base.

Figure 2

The "Help" Command

```
 G I P S Y - UNIVERSITY OF OKLAHOMA

help
-------
                        GIPSY COMMANDS

MESSAGE - SENDS A MESSAGE FROM THE TERMINAL TO THE
          CONSOLE OPERATOR AND WAITS FOR A REPLY

DEFINE  - DISPLAYS ALL FORMS IN THE DICTIONARY

FORM    - IDENTIFIES A SPECIFIC DATA BASE
          ISSUED AT THE BEGINNING OF A GIPSY RUN

SELECT  - INITIATES A SEARCH ON THE TOTAL DATA BASE
          AND SELECTS A SUBSET

ITERATE - INITIATES A SEARCH ON A SELECTED SUBSET
          OF THE DATA BASE

LIST    - LISTS RECORD ITEMS FROM A SELECTED SUBSET
          ON THE TERMINAL OR PRINTER

PRINT   - PRINTS ENTIRE RECORD(S) FROM A SELECTED
          SUBSET ON THE TERMINAL OR PRINTER

COPY    - COPIES DATA FROM A SELECTED SUBSET ONTO
          A FIXED FIELD FILE

SORT    - SORTS ON SELECTED ITEMS OF A SUBSET

SUM     - PROVIDES AN OCCURANCE COUNT, AVERAGE, MAXIMUM AND
          MINIMUM FOR NUMERIC ITEMS OF A SELECTED SUBSET

BACK    - SELECTS A PREVIOUS SUBSET OF THE SELECTED RECORD
          FILE (SRF) TO BE THE CURRENT MOST ACTIVE SRF

DUMP    - COPIES RECORDS FROM A SELECTED RECORD FILE ONTO
          A SEQUENTIAL FILE IN DUMP/RESTORE UTILITY FORMAT

END     - TERMINATES JOB EXECUTION

WHEN ENTERING COMMANDS TYPE '???' FOR DETAILED EXPLANATIONS
ENTER A COMMAND.....
```

Both a copy of the procedure for securing the dictionary and the dictionary itself for the ERIC records are presented in Figure 3. The labels in the left-hand column are mnemonic codes for searching the several sections of the ERIC records. Each label can be related to a specific part of a record. A sample of an ERIC record is presented in Figure 4 to permit the reader to associate the labels in the dictionary with the entries in the record. The format of the record has been altered slightly from that used in RIE.

Experience has shown that when information about specific topics or programs is desired, the labels used most frequently in the search will be the titles, abstracts and descriptors. A primary goal of the search is the retrieval of the accession number which permits the user to locate the microfiche of the total document.

The user should note that the command "help" is used before a data base is selected, whereas the "define" command requires that a data base must be specified first. "Help" relates to the general operation of GIPSY; "define" is specific to a data file, such as ERIC.

"???" A third source of assistance for the operator is the command "???." It is entered after the "select" command and will cause the computer to print suggestions on a number of strategies or tactics available to the searcher. A copy of the response to "???" is presented in Figure 5.

Operating the System

The operation of the GIPSY/ERIC system can be described in a sequence of five steps: initiation, input, manipulation, output, termination. Each step requires certain procedures which must be followed with some precision.

Figure 3

The "Define" Command

```
 G I P S Y - UNIVERSITY OF OKLAHOMA

define
-------

ERIC

ITEMS FOR FORM?
yes
ENTER FORM
eric
TERMINAL OR PRINTER?
terminal

LABEL                DEFINITION

  ID          ACCESSION NUMBER -
  CID         CLEARINGHOUSE ACCESSION NUMBER -
  PGMAREA     PROGRAM AREA -
  PUBDATE     PUBLICATION DATE -
  INCODE      INSTITUTION CODE -
  SPCODE      SPONSORING AGENCY CODE -
  EDRS        EDRS PRICE -
  DESNOTE     DESCRIPTIVE NOTE -
  ISSUE       ISSUE -
  REPORT      REPORT NUMBER -
  CONTR       CONTRACT NUMBER -
  GRANT       GRANT NUMBER -
  BUREAU      BUREAU NUMBER -
  AVAIL       AVAILABILITY -
  JRNL        JOURNAL CITATION -
  INSTIT      INSTITUTION NAME -
  SPONSOR     SPONSORING AGENCY NAME -
  TITLE       TITLE -
  AUTHOR      AUTHOR -
  ABSTR       ABSTRACT -
  DESCRIP     DESCRIPTOR -
  IDENT       IDENTIFIER -

ANOTHER FORM?
no
```

Figure 4

An ERIC Record

ACCESSION NUMBER - ED003552

CLEARINGHOUSE ACCESSION NUMBER -

56
PUBLICATION DATE - 64
INSTITUTION CODE - MGG30750
EDRS PRICE - EDRS PRICE MF-$0.36 HC-$8.16
DESCRIPTIVE NOTE - 204P.
ISSUE - DH6667
REPORT NUMBER - NDEA-VITA-191
INSTITUTION NAME - HARVARD UNIV., CAMBRIDGE, MASS.

TITLE - AN ANALYSIS OF THE BEHAVIORAL PROCESSES INVOLVED IN SELF-INSTRUCTION WITH TEACHING MACHINES.

AUTHOR - HOLLAND, JAMES G.; SKINNER, B.F.

ABSTRACT - THIS COLLECTION OF PAPERS CONSTITUTES THE FINAL REPORT OF A PROJECT DEVOTED TO AN ANALYSIS OF THE BEHAVIORAL PROCESSES UNDERLYING PROGRAMED INSTRUCTION. THE PAPERS ARE GROUPED UNDER THREE HEADINGS--(1) "PROGRAMING RESEARCH," (2) "BASIC SKILLS--RATIONALE AND PROCEDURE," AND (3) "BASIC SKILLS--SPECIFIC SKILLS." THE SUMMARY WRITTEN BY THE PRINCIPAL INVESTIGATOR IS GIVEN IN A PAPER TITLED "REFLECTIONS ON A DECADE OF TEACHING MACHINES." MANY OF THE PAPERS HAVE BEEN PUBLISHED IN PROFESSIONAL JOURNALS. (AL)

DESCRIPTOR - *AUTOINSTRUCTIONAL AIDS; *BEHAVIOR PATTERNS; LINEAR PROGRAMING; PROGRAMED INSTRUCTION; *TEACHING MACHINES; *VERBAL LEARNING

IDENTIFIER - CAMBRIDGE; MASSACHUSETTS

Figure 5

The "???" Command

```
select
-------
A. ???

   ENTER SEARCH CRITERIA

   PARAMETER        EXPLANATION
   ---------        -----------

   LABEL            NAME DEFINING AN ENTRY IN A RECORD
   DESCRIPTION      ANY COMBINATION OF LETTERS OR NUMBERS
   NUMBER           DECIMAL NUMBER
   WI               WITHIN
   EQ               EQUAL
   GT               GREATER THAN
   LT               LESS THAN

   POSSIBLE COMBINATIONS:

   A. LABEL
   B. LABEL<DESCRIPTION>
   C. LABEL<DESCRIPTION> THRU <DESCRIPTION>
   D. LABEL NUMBER THRU NUMBER
            EQ
   E. LABEL GT NUMBER
            LT

            EQ
   F. LABEL GT LABEL
            LT
   G. LABEL<DESCRIPTION> WI NUMBER <DESCRIPTION>
            TWO DESCRIPTIONS OCCUR WITHIN  X WORDS OF EACH OTHER
   H. LABEL <DESCRIPTION> NUMBER
            DESCRIPTION OCCURS AT LEAST THE SPECIFIED NUMBER OF TIMES

   THE LOGIC ENTRY REQUIRES THE SEARCH CRITERIA TO BE ASSOCIATED BY
   (AND, OR, NOT) TO SPECIFY THE CONDITION FOR RECORD SELECTION

   EXAMPLE:
            SELECT ALL MALES BETWEEN THE AGES OF
            45 AND 55 WHO ARE IN SUPERVISORY
            POSITIONS AND EARN A SALARY GREATER
            THAN $10000

   A. MALE
   B. AGE 45 THRU 55
   C. TITLE<SUPERVISOR>
   D. SALARY GT 10000.00

   LOGIC A AND B AND (C OR D)

   ENTER SEARCH CRITERIA.....
```

Initiation

The initiation of the operation of the GIPSY/ERIC system is commonly referred to as "bringing up" the system or as a "hookup" procedure. The terminal is connected to the computer through the use of a regular office telephone with the aid of an acoustic coupling device. The device is a small box into which the telephone receiver fits. Signals are transmitted from the remote access unit through the coupler to the telephone line; at the other end another coupler provides access to the computer and a data file. Signals can be returned to the keyboard from the computer using the same system.

The operator at the terminal must observe three signal lights. A blue light goes on when the system is ready for operation; however, the operator cannot proceed until the green light comes on. The red light indicates trouble on the line. The lights are controlled by the computer and inform the user about the operation of the system. Thus, although the user dominates the operation of the system, he can do so only when the machine permits him. For example, after he has requested a search, the blue light will indicate that the search is being conducted. He cannot interrupt the search nor can he initiate additional inquiries until the system is returned to his control and the green light is on. With a green light, he has complete control of the system.

Several rules must be followed in initiating a search. The operator should be sure that the switch on the acoustic coupler is in the "off" position before placing the telephone receiver into the coupler. The procedure insures that no signals from the computer will be lost before the operator is ready to begin his search. The terminal switch must be

in the "on" position during initiation. With the terminal switch in the "on" position and the coupler switch in the "off" position, the operator dials the appropriate telephone number at the computing center. When contact is made he receives a high-pitched signal. He then places the receiver in the coupler and moves the coupler switch to "on." The terminal is now ready for operation and the machine prints the identifying citation:

G I P S Y - UNIVERSITY OF OKLAHOMA 2:44 P.M. TUESDAY, FEBRUARY 2, 1971

Input

The input for the search is analogous to a program deck in other computer operations. It is quite different from the usual program in that it is not pre-prepared and can be revised at the terminal.

In operating the system the first action is the selection of a specific data base by the use of the command "form." In GIPSY both the RIE and the CIJE records are filed under form ERIC.

The records are made available when the code word "eric" is entered after the command "form" has been acknowledged by the machine. Acknowledgment is received at the terminal in the form of seven dashes beneath the command. When the ERIC records have been identified and are ready for search, the carriage moves the paper down two lines and is prepared to accept a command from the ERIC dictionary.

The command "select" is entered next. This indicates that the terminal operator wishes to initiate a search of the ERIC records for specific terms or combinations of terms. The machine acknowledges the "select" command with seven dashes. It then provides a label for the

first variable by printing a capital letter "A." The terminal operator then switches from system operating commands to any of the labels from the ERIC dictionary. Although there is a real limit on the number of labels and terms which may be entered prior to beginning a search, because each variable must be labeled with a different letter, the practical limit is the amount of time required by the complexity of the search.

Manipulation

The GIPSY commands permit the manipulations of the records before they are printed. The manipulation commands allow the operator to organize the records in a variety of patterns.

Each command is sent to the computer by returning the carriage. After the operation has been performed, the user receives a report on the result of the manipulation. Usually it will take the form of a count of the number of documents which were identified through the manipulation. In this way the user can browse through the base applying all of his professional knowledge to explore its content.

The "iterate," "sort," "sum" and "back" commands are examples of the manipulations available to the user. The command "iterate" allows the operator to return to subsets previously established in order to re-search the records in that set perhaps for other, additional information or to select other variables. If he wishes to return to the total record set, he would use the "select" command.

Quite often as the user reduces the size of subsets he becomes aware of certain characteristics of the data; and, in order to insure that he will not have missed significant information, he will return to

an earlier data set and pursue the search using additional or different variables. For example, a search was conducted for information relating to records on defacto segregation. In the initial search "defacto" was entered as a single word. The user, upon receiving the count, suspected that he was not getting all of the records. He returned to the base and entered the term as two words (de facto) and as a hyphenated word (de-facto). He learned from the return that all three forms are used interchangeably in the ERIC records. He must keep information such as this in mind if future searches are to be different.

After a search is completed, the term "ITERATE?" (with the question mark) is typed by the computer to ask the operator if he wishes to continue searching within the current subset. It provides the operator with three options:

1. Yes, which may be abbreviated "y" or spelled out. The machine then makes the current subset available for further exploration and reduction.

2. No, which may be abbreviated "n" or spelled out. The response indicates that the user is either satisfied with what he has received or he wishes to pursue an alternate strategy.

3. Yes, followed by a comma and number (y,3). This would indicate that the user wishes to enter a subset other than the one that has just been established. The number which he enters identifies the subset with which he wishes to work. In the example, he would return to subset three. All subsets which have followed the one to which he is returning will be erased. The program feature assumes that the user has finished with the records selected for subsequent sets; return to an earlier set indicates

a new direction, if not a completely new search.

The "iterate" command may also be used after the operator has selected a set of records and asked for the information to be printed. He may elect to go back to a previous subset to continue his search of the same topic using different variables or a different strategy. The procedure is to type the word "iterate" at the terminal. The computer will ask for identification of the desired subset.

The command "back" is also available to permit the operator to return to a previously formed subset. If the "back" command is used, the computer will request that the subset be specified. The operator must enter the "iterate" command to activate the system for further questions. More often, the "back" command will be followed by a processing command for further manipulation of the subsets of the record file selected by the operator.

The processing commands include "sort" and "sum." The "sort" command allows the operator, or the user, to organize the information he has located in several ways. Using the "sort" command the user may have the abstracts or other information placed in numerical order, either ascending or descending, or organized in alphabetical order. With the ERIC base there are several ways that "sort" may be used to place the information in some useful order. One useful sort would be select on the basis of the ED number provided on each ERIC abstract. Another method of sorting the selected items would be through the use of dates of publication. The user may want to arrange the records in alphabetical order by author. With the ERIC file, the latter can be a dangerous procedure since not all records contain an author's name. A similar problem may be faced if

the user wishes to arrange the records by the source of funding for the project described in the report; not all records describe funded projects.

The "sum" function, which is also a processing command with GIPSY, allows the operator to obtain sums, averages and occurrences of counts in the information. Since the ERIC records are primarily narrative, the "sum" command is of little use with them. However, a user might be interested in the number of document pages represented by the records. he has selected. The command is very helpful with other data, such as survey research.

Occasionally the user will make a typographical error or he will change his mind after he has started to type a command. He has three procedures available to correct an error or to indicate that he desires a different procedure. The typographical error can be corrected simply by backspacing and correcting the entry. It is most useful when only a few additional characters or spaces have been entered. It may be more effective to remove the entire line. He can do so by returning the carriage. The action will cause him to receive an error message indicating an invalid entry. The machine will move the paper down one line and the user may begin again. Finally, he may erase an entire line by depressing the interrupt key. Again, the machine will return the carriage and he may proceed as if the previous line had not been entered.

In posing his inquiry the user has several options. He may use total words, in which case the word will have a blank space (␣) between its beginning and the first delimiter and another space between its end and the closing delimiter: <␣school␣>. He may choose to search only suffixes. In this case he would enter the characters with no blank

between the first delimiter and the first letter but with the blank following the string of characters and before the second delimiter: ⟨school∅⟩. Similarly, he may search for prefixes by inserting a blank before the string of characters but having the delimiter follow the last character immediately: ⟨∅school⟩. The user's professional knowledge can permit him to search more efficiently by using only parts of words rather than the entire string of symbols. The system also provides him with the ability to search simultaneously for a number of related terms. He could enter, for example, ⟨∅teach⟩. The search would select all records which contained "teacher," "teachers," "teaching," as well as "teach" and "teaches."

Once the operator has entered the variables, the terms to be selected by the search, he is ready to activate the search. The procedure is to return the carriage after the computer has entered the letter for another variable. In response, the computer will type the word "LOGIC." After the operator has received control of the machine he specifies the relationship that he wants among the variables. The logic statement is limited to the letters used to label the variables. These can be connected using Boolean logic to answer the question that the user is posing. Some principles of Boolean logic will be presented in Part IV. The logic statement is sent to the machine in the same manner that the variables were transmitted, simply by returning the carriage.

The computer then types "LOGIC" again and the operator may assemble the variables into a different order. This is known as the multiple logic sequence. The system is presently capable of accepting up to nine additional logic statements. The multiple logic statements will not

produce a subset; only the initial logic statement establishes a subset.

The value of the multiple logic statement is in discerning the number of records which would have been available had the operator searched with a different logic. The computer will give counts on each of the logic statements which are entered. The strategies for the multiple logic statements will be discussed in the next section along with the general logic statement.

After the operator has entered the logic statements that he desires, he will return the carriage when the computer types "LOGIC" again. He will then receive the message "SEARCH BEGINNING." At this point the computer takes over and the operator has a short waiting period while the computer goes through the records to find the abstracts which contain the variables in the logical order that the operator has requested. Once the computer has located the documents which fit the logic, it will print out a numeric analysis of the logic statements satisfied. The computer will report the number of records searched and the number of records that the computer has selected which fit the first logic statement. The computer will also report the number of records containing the variables which were requested.

Output

There are four output commands with GIPSY. These allow the operator either to receive printed copy or to transfer the selected files to other devices, such as disc packs or tapes, for future processing.

Upon completion of a search the operator may ask for printed copy of the records selected. He has two commands available for this purpose. If the operator desires the entire content of each record in the subset,

he would use the command "print." The command will reproduce the entire entry from the subset. He must, however, indicate whether the records are to be printed at the terminal or on the high-speed printer at the computing center. Printing at the terminal is a slow process. The terminal is capable of printing only 100 characters a minute. In contrast, the printer at the computing center can produce 1,000 lines of print in a minute. Therefore, for a subset containing more than three or four records, it is more efficient to have them produced through the high-speed printer. For the records to be printed at the computer center the user at the terminal must enter a label which will be used to identify his records and to provide an address so they can be delivered to him.

If he elects to print at the terminal, the machine will present 20 lines of copy and stop. He may request an additional 20 lines by simply returning the carriage. In contrast, after examining the material in the first 20 lines he may use the interrupt key to discontinue production of the terminal. He can then use the "iterate" command, specify the subset and have the information printed at the high-speed printer. He thus has the ability to review the content, or parts of the records, before ordering the final output.

If the total record is not desired or needed, the user may give the command "list" to the computer. The "list" command allows the operator to select the parts of the record that he desires for printing. The command also allows the items to be printed at either the terminal or the high-speed printer.

The ability to review sets of 20 lines presents an alternate strategy. The user may elect to review the titles of the abstracts or

perhaps the authors at the terminal before having the entire records printed at the computing center. He may do this by using first the "list" command and requesting output at the terminal and subsequently using the "print" or another "list" command for the entire output at the computing center. Also, after reviewing a partial output by using the "list" command, he may elect to reduce the set further before printing. He can do so by initiating an "iterate" command.

Should the operator not desire to have the information printed by the computer and wish to make other uses of the records, he may enter one of the two commands: "copy" or "dump." The command "copy" causes the computer to copy the selected subset of information onto a fixed field file so that further manipulations using other programs may be carried out. If he is working with statistical data, he may copy the material and then request further processing from the Standard Statistical Package (SSP) or perhaps the Biomedical Computer Programs (BMD).

The "dump" command also copies the data from the selected record onto another file. It allows the user to return to the subset at a future date to continue the search. The "dump" command is especially useful in reducing the total size of subsets for terminal operations. It can be used through batch processing, which will be covered in a later section of this monograph.

A command that can be used at any point in the program and one that has a variety of uses is "message." The "message" command permits the user of the remote terminal to communicate with the operator of the computer at the center. He may wish to alter his procedure in such a fashion that it requires a physical manipulation of the hardware at the center.

He may, for example, wish to have a set of records, which he has previously selected by the use of batch process, to be made available for further investigation at the terminal. Also, he may wish to give special instructions, such as printing multiple copies of the materials which he has selected. The machine operator can also communicate with the user at the terminal through the system. For example, the operator of the GIPSY/ERIC system requires a pre-scheduled time period. The machine operator may wish to advise the user that he is nearing the end of his run time. Such a warning can prevent the loss of a carefully developed set of records when the time period expires.

Termination

When the user has completed his run, the final operation is to terminate the contact with the computer. Standard usage suggests that a message to the operator may be helpful. In some instances, the user may wish to specify multiple copies of the output. This requires the assistance of the operator at the computing center. He may have special requirements for delivery or pickup of the materials retrieved from the system. Again, this can be done using the "message" command, or it can be done on an alternate telephone line.

The operation of the system is terminated by use of the command "end." When that command has been received, the machine will respond with seven dashes. It then produces its standard sign-off routine:

G I P S Y - UNIVERSITY OF OKLAHOMA 4:01 P.M. TUESDAY, FEBRUARY 2, 1971

Once the system has been signed off, it is necessary to shut down the equipment. This requires three steps:

1. Move the switch on the acoustic coupler to the "off" position.
2. Remove the phone from the coupler and replace it in its cradle.
3. Turn off the remote access unit.

It should be noted that the materials selected are retained in the machine in temporary storage until the system is signed off. At that time it is printed. For this reason it is possible for the user to wait until the end of his run to specify special instructions, such as the use of two-part paper for the retrieved records.

Recapitulation:

The Search Process

A summary and review of the search process is presented in the flowchart in Figure 6.

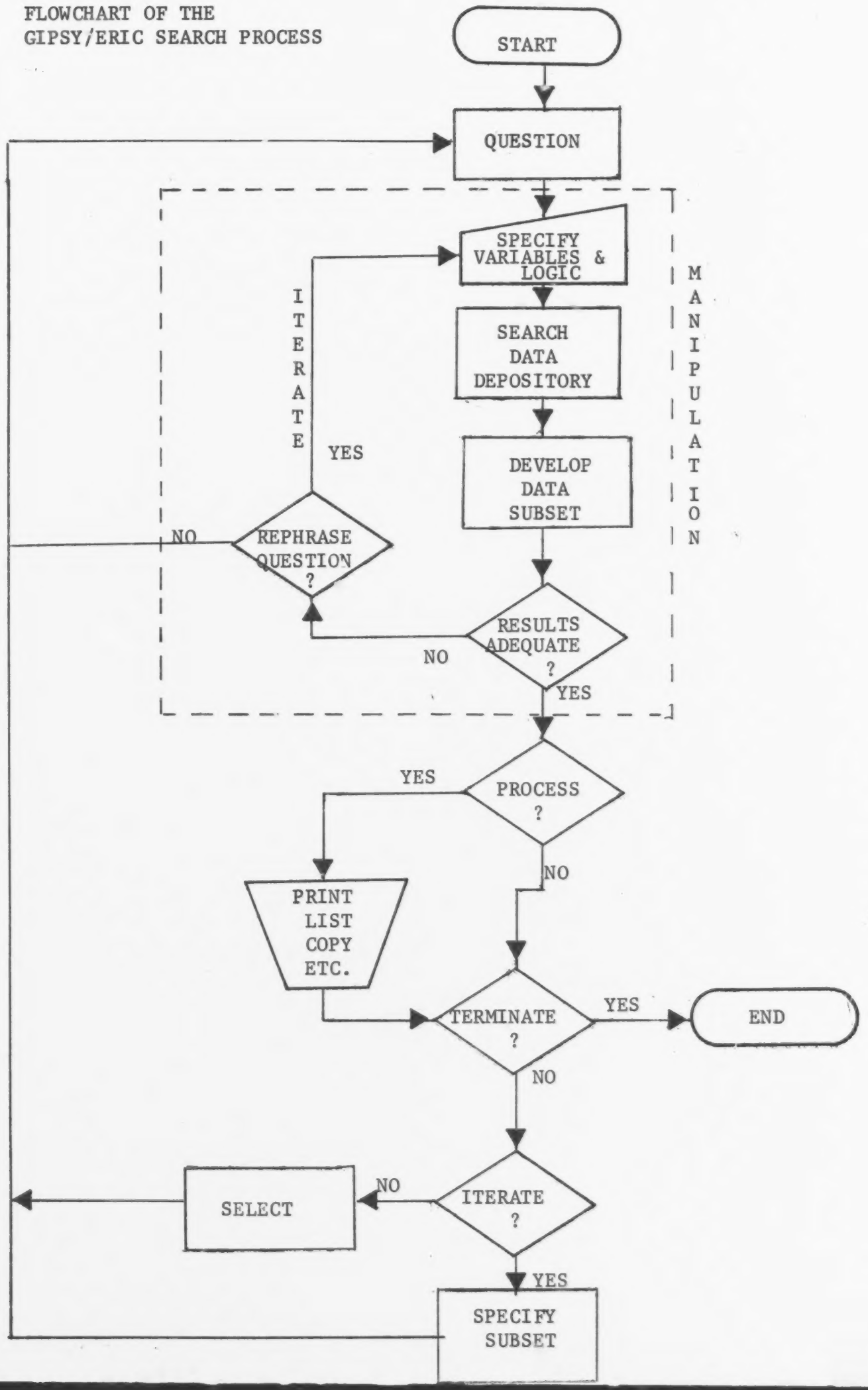
Once the system has been initiated (START), the search begins with the identification of a pertinent question. In its initial form the question will be stated in standard English. Subsequently, key terms must be specified. The second step is indicated on the flowchart by the symbol for an on-line keyboard. It indicates that the user must specify and enter the variables and the logic. The system is under the direct and immediate control of the user. The search of the data depository begins when the logic statement has been entered. The search process will establish a subset of data and report on its composition to the user. Based upon the information in the report, the user must evaluate the adequacy of the subset. If in his judgment the results do not satisfy his question, he may elect to iterate. He can rephrase the question within the existing subset of data and seek further specification within the manipulation loop.

If in his opinion the results do not warrant further exploration, he can leave the loop and begin a new approach to the total data file by formulating a new question and developing different variables and logic statements. If the report on results is adequate, he has several options.

If he wants to examine the records that he has identified, he may elect to process them through one of the output commands ("print," "list," etc.). In some cases the information gained from the record count will be adequate for his purposes. In this instance he may choose to terminate the search by entering the "end" command, or he may wish to begin exploring a different question by using the "select" command. Also, he could return to an earlier subset or to the total data base.

Following the output processing, he must decide whether or not to continue his search. If he does not wish to continue, he will then proceed to terminate the operation. If he does wish to continue, he has a choice of specifying a subset or returning to the total data base.

FLOWCHART OF THE
GIPSY/ERIC SEARCH PROCESS



PART IV

BOOLEAN LOGIC:

The Basis for Information Manipulation

The operation of the computerized information system requires three varieties of knowledge. As was pointed out earlier, the most important for the user is his professional background. The second significant class of knowledge with which he must work concerns the characteristics of the data base. Finally, he must have certain logical tools to specify the manipulation of the data according to his professional acumen. The most widely used process for the latter activity is the application of Boolean logic.

Boolean logic was devised over a century and a half ago by George Boole, an English mathematician and schoolmaster (Sciences, 1965). His approach to logic was built upon four significant points:

1. The use of only two symbols, zero and one. Because of this propitious choice, his approach was readily adapted to the functioning of the electronic computer.

2. The specification of homogeneous classes or systems for classification.

3. The use of symbols to denote class attributes.

4. The use of symbols to denote logical operations similar to those of algebra. His work is often cited as a major advance in science. It provided a break with the historical tradition that decision making was primarily the domain of ethics and theology. His work placed it clearly

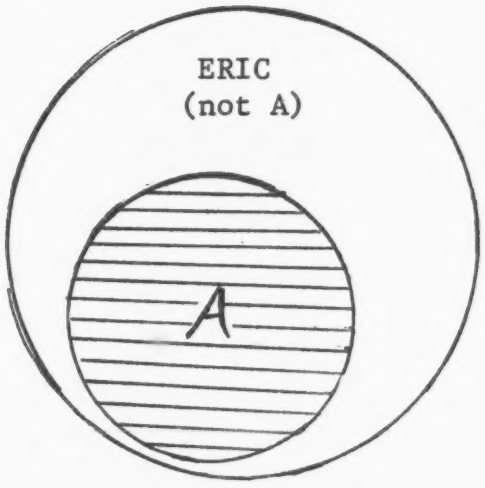
in the realm of mathematics and science.

The logical operators for the GIPSY/ERIC information system, are based upon a dichotomy of existence or nonexistence of a particular set of symbols. The symbol set need not be a standard word. In many instances, as was pointed out in the section on the manipulation of information, only a fragment of a word will be used. In other instances, special symbols will be attached in a data file as an aid to indexing. In the ERIC descriptors, for example, an asterisk is appended to some words indicating that they are considered major terms in the classification system. The user can captalize upon this peculiarity of the data base. For some searches it will be important to precede a key word with an asterisk. If a space should be entered instead, the term would not be identified and a valuable record could be lost.

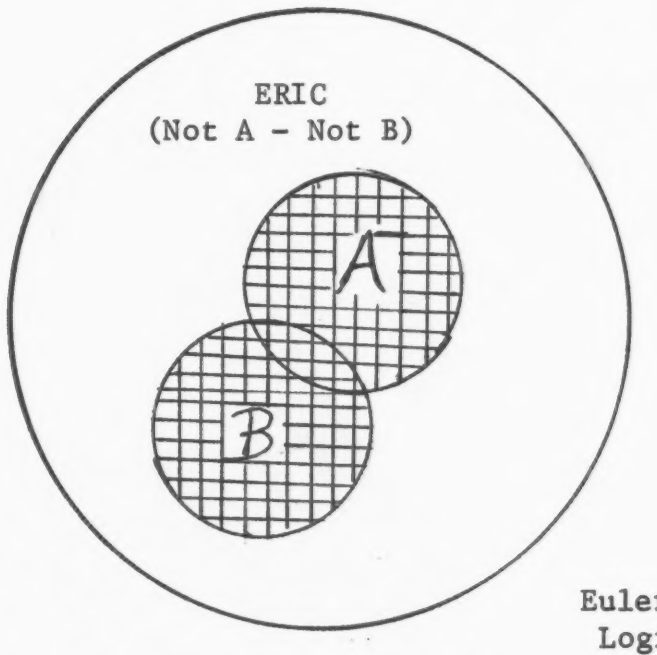
Boolean Operators

Boolean logic presents the user of the system with four ways of combining the variables in which he is interested. As was pointed out in the section on manipulation of records, it is possible in GIPSY to specify a sequence of logical combinations of variables and thus gain additional information about the topic being studied. The four logical operators can be combined in a number of ways to produce a relatively simple or highly complex logical structure:

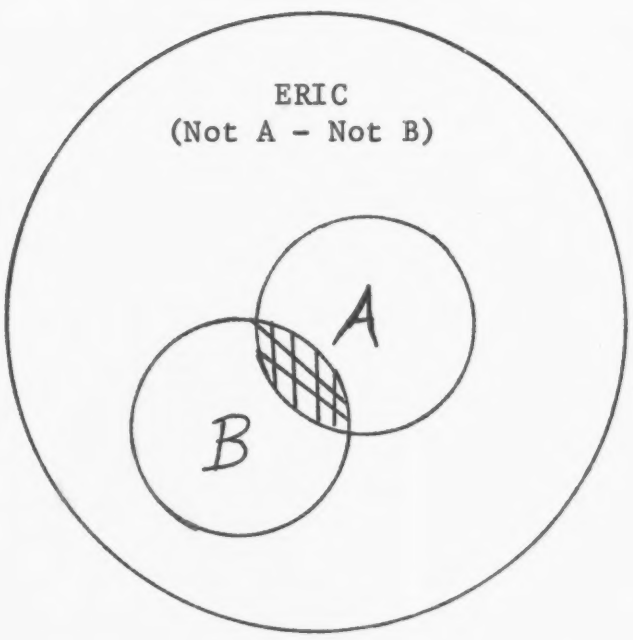
1. Logic A. The function of the Logic A operator is to remove from the file all those records containing Variable A and establish them as a separate subset. The shaded area of the Euler Diagram I represents visually the functioning of this logical operator.



Euler Diagram I
Logic A



Euler Diagram II
Logic A or B



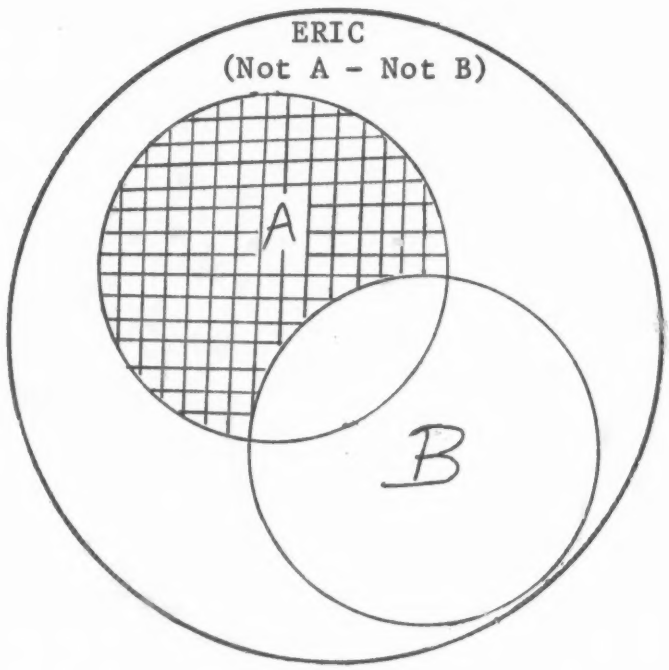
Euler Diagram III
Logic A and B

2. Logic A or B. Two variables can be combined into a subset, using the Logic A or B mode. Additional variables can, of course, be added to the string. The shaded area of Euler Diagram II represents the document subset required to satisfy the Logic A or B operator. It does not matter whether the record contains one or more instances of either variable, but it is required that somewhere in the document, one of them must appear. A document may contain only Variable A or only Variable B; other documents will contain both A and B and perhaps be repeated several times. The first operator, Logic A, is a special case of the Logic A or B mode.

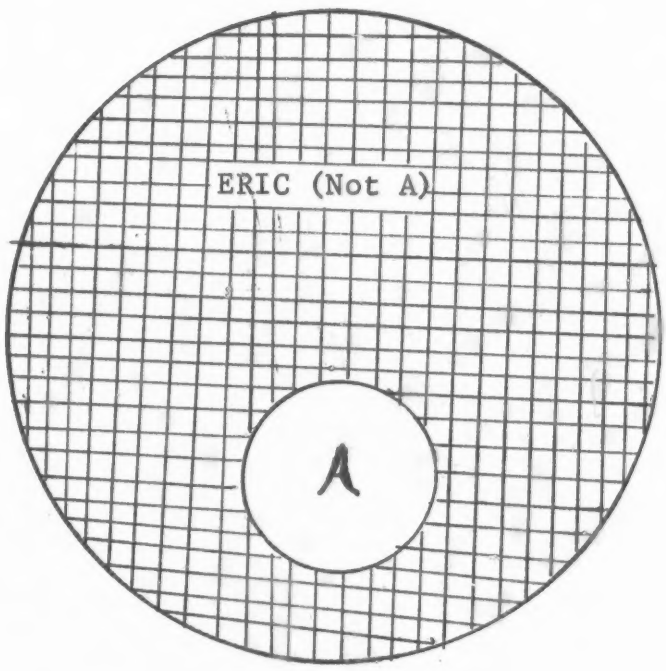
3. Logic A and B. Sometimes the user of the system will require that both A and B be present in each document. The Logic A and B operator establishes the criterion that both variables must be present. Euler Diagram III provides a visual interpretation of the logical operator. The subset which results from the operation will not include documents that contain A but do not contain B, nor would it include documents containing B but not A. Both variables must be present.

Just as with the A or B operator, additional variables may be added to the string. For example, one user could request A or B or C or D...; and another A and B and C and D.... Usually, adding variables with the "or" operator will increase the size of the set while with the "and" operator, set size will decrease rapidly.

4. Logic A and not B. Just as it is possible to require both variables to be present in one document, it is possible to require that one be present and one be absent. The operator Logic A and not B may be used to establish such a subset. The result of the operator is illustrated in Euler Diagram IV. All documents containing B will be excluded



Euler Diagram IV
Logic A and not B



Euler Diagram V
Logic Not A

from the subset, even though it may appear within some of them. Of course, it would be equally feasible to reverse the order and require B to be present, but not A. Again, more than two variables may be used.

A special case of the A and not B logic occurs when all of the documents within a set are desired except those containing a given term or set of terms. The logic reduces to "not" A. Euler Diagram V presents a description of the results of the operation. All of the documents containing A will be removed from the subsets. The user may, for example, want to identify documents pertaining to evaluation or accreditation of four-year colleges, but specifically not any of the records containing information about junior colleges or universities. A major variable for the search is "college" (A). A second variable would be "evaluation" (B) or "accreditation" (C). Finally, the variables "junior" (D) or "university" (E) must be entered. A complex logic statement can be written which would identify reports about the evaluation or accreditation of colleges but eliminate junior colleges and universities. Since the concern is with autonomous four-year colleges and not with colleges located within a university, it will be necessary to eliminate from the records those reports of colleges as subunits of a university. For example, the user was not interested in reference to the College of Education of the University of Oklahoma. The logic statement could be written:

(A and (B or C)) and not (D or E).

The parentheses are used to clarify the operational sequence. Extremely complex logic statements can be written. Usually, the initial stages of a search will use a relatively simple logic. Most commonly,

the A or B logic will be used in order to provide a maximum number of records in a subset which can be revisited for sequential searches. As the number of records is reduced, run time becomes sufficiently brief that the more complex forms of logic become reasonable.

Only those variables which have been identified by a letter may be entered into a logic statement. It is not necessary to enter all variables in the logic statement. The machine will return a count on variables identified, even though they are not included in the logic statement. Obviously, the number of variables which can be entered is limited by the 26 letters which can be used to label variables. General practice suggests that, while the machine is capable of handling variables nested five or six deep, the human mind can grasp the meaning of only about three levels.

In entering logic statements into GIPSY, several alternate approaches may be used. A "+" may be substituted for the "or" and a "*" may be substituted for the "and." Either form may be used and both may be used within one logic statement. Also, it is not necessary to separate words and symbols by spaces. The spaces are acceptable, and will cause no difficulty, but they are not necessary.

Earlier, it was pointed out that parentheses may be used to clarify logic statements. In some instances, they will be required to get the meaning desired by the systems user. Consider the differences in the three statements:

(A*B)+C

A*B+C

A*(B+C)

The first requires that both A and B must be present or C alone may be present in order for a document to be identified. The next two require that Variable A must be present in a record along with either Variable B or C. When one set of parentheses is enclosed by another, the innermost takes precedence and is performed first. By a sequence of nesting, extremely complex searches of a data file can be arranged.

The use of the logic statement is perhaps the most difficult ability the user must acquire to run the GIPSY/ERIC system. Most users have found, that while initially they can think only in terms of relatively simple combinations, as they gain experience with the system, they become increasingly able to use complex logic to describe the combinations of variables which they desire.

PART V

Sample Searches

The following section presents two examples of searches of the ERIC data base which were made using the GIPSY system. Copies of the terminal printouts are enclosed and descriptive comments on the process are presented. One search is relatively simple and the other complex.

A Simple Search

A request for a search of the ERIC data base was received from a college which was proposing to develop a program to assist underprivileged Negro students to profit from a college experience. The research group submitted eleven specific questions. Figure 7 presents these questions.

An initial decision in conducting a search concerned the identification of a sequence of terms to select a set of records which contain useful information. The decision in this search was to begin by selecting from the total data base all the records concerned with universities and colleges. Figure 8 presents the first sequence in the search as it was conducted. Note the heading "GIPSY--UNIVERSITY OF OKLAHOMA," which indicates that the system is operating and the user may proceed.

The input is in lower case letters; the system responds with capital letters. The user can thereby reconstruct his search procedure at a later time and evaluate his strategies. The request for "form" received a response of seven dashes. The user then requested the ERIC data base. Subsequently, he indicated that he wished to select certain records.

FIGURE 7

Initial Questions for a GIPSY/ERIC Search

1. Research related to criteria used in admitting Negro adolescents into American colleges and universities, how to enroll them, types of criteria used.

2. Longitudinal research describing those admission criteria which are best predictors of success or failure in academic work.

3. Remedial programs administered by colleges or contracting agencies to improve chances of success in college for black students.

4. Research related to attitudes of black adolescents towards college, higher education and white instructors.

5. Personality characteristics of the urban Negro adolescent: his IQ, need for achievement, source of motivation.

6. Academic achievement of Negro adolescents in Dallas/Ft. Worth area (or southern/southwestern urban areas); level of verbal skills, understanding of spatial relationships, learning processes.

7. Research on the effectiveness of tutorial programs (one to one teaching), for disadvantaged Negro youth entering college.

8. Attitudes of college faculty to admission of disadvantaged youth who do not meet standard admission criteria, teaching remedial courses, risk students in class.

9. Inter-cultural sensitivity training programs for students and faculty (black and white).

10. Interaction analysis of tutor-student relationship--white/white, white/black; black/white; black/black.

11. Models for facilitating the learning processes of disadvantaged Negro adolescents.

Figure 8

First Sequence of a Simple Search

```
G I P S Y - UNIVERSITY OF OKLAHOMA  MONDAY  JUNE 29, 1970

form
-------
eric

select
-------
A. title <university>
       TITLE -
B. abstr $
       ABSTRACT -
C. $ <college>
D. title $
       TITLE -
E.
LOGIC a or b or c or d
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED   27026
 SELECTED    8063      SUBSET 1
 VARIABLES SATISFIED
     A          840
     B         3898
     C         5022
     D         2017

ITERATE?
yes
```

It is possible to use either the entire word "select" or merely the first three letters.(sel).

Titles and abstracts were searched for the terms "university" and "college." The symbol "$" serves as a ditto mark; it should be noted that it can be used either to repeat a label or to repeat a content within the delimiters which indicates the term or terms the system should search for.

Four items were included in the initial search. Note that the term "university" would not include "universities." The delimiter immediately after the final letter in "college" would permit its use as a prefix so that the plural, "colleges," would also be identified. When a request came from the system to enter Item E, the user ended the input sequence by returning the carriage. The machine then indicated it would accept a logic statement. The logic statement employed the simple "or" logic; the user wanted to see how many times each of the terms, "university," "college" or "colleges" appeared in either abstracts or titles. When the logic message was sent, the machine responded with the word "SEARCH" and then asked if a second logic statement was desired. By returning the carriage without entering a second logic statement, the search began.

When the search had been completed, the system established the selected records as "SUBSET 1" and reported to the user the total number of records searched, the number of records which satisfied the logic and the number which satisfied each of the variables. The user knew after the first run that the term "university," (Variable A) appeared in 840 titles; however, as Variable B it was found in 3,898 abstracts. Following the return message, the system requested instruction on whether

the user desired to further reduce subset one by iteration or whether he desired to return to the total base. He elected to proceed and responded to the "ITERATE?" request with a "yes." He could have used the abbreviation "y."

For the second search the user decided to narrow the set of those records concerned with culturally deprived or culturally disadvantaged students. Figure 9 presents a record of the run. Once again he searched only titles and abstracts. Note that he used the prefix "cultur" which would identify for him such words as "cultural," "culturally," "culture," etc. Similarly, in Variable E he asked for the prefix "disadvant." This would select for him words such as "disadvantaged," "disadvantage," "disadvantages" or "disadvantageous." Since the focus of the question was on the culturally different child, the word "culture" would be acceptable in either the titles or the abstracts but must be linked to some form of "deprived" or "disadvantaged" which again could be located in either the title or the abstract. Thus, the logic statement showed a rudimentary form of parenthetical logic: Two sets of symbols, coded by the letter designating the variable, were linked together with the "and" logic. There had to be one element from each set present in a record before it would be selected. The procedure resulted in "SUBSET 2" which contained only 145 records. An examination of the report on variables satisfied revealed that there was far more information contained in the abstracts than in the titles. The difference is quite reasonable since the abstracts are considerably longer than the titles.

The decision was made to iterate and further reduce the number of

Figure 9

Second Sequence of a Simple Search

```
A. title < cultur>
        TITLE -
B. abstr $
        ABSTRACT -
C. $ <deprived>
D. title $
        TITLE -
E. $ < disadvant>
F. abstr $
        ABSTRACT -
G.
LOGIC (a or b) and (c or d or e or f)
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED     8063
 SELECTED      145      SUBSET 2
 VARIABLES SATISFIED
      A          64
      B         574
      C          74
      D           9
      E         111
      F         387

ITERATE?
yes
```

records in subset two. In the third iteration (see Figure 10) the terms "urban" and "Negro" were requested. Once again, titles and abstracts were searched. Although the planners of the project had used the term "black" and "Negro" as synonyms, the user's knowledge of the data base suggested that because of the age of the records, there would be no point in using the word "black" as a synonym for "Negro"; it is a much more recent designation.

Note that Variable C was entered three times. The first time was a typing error. In the second the "$" sign requested a duplication of the term which had been entered under Title. Thus, Items A and C would have been identical. The user was able to interrupt the system and correct the error by requesting the term "urban" in the abstracts.

Note that the form of the logic used in the search was changed. Rather than using the word "or," the user substituted a plus sign (+).

Subset three reduced the number of qualified records to 48. The difference in counts between the content of titles and the content of abstracts was dramatic in this run also. The decision was made to iterate again. The program is presented in Figure 11. The thought had occurred to the user that the requestors were not concerned with programs in junior colleges, only those in four-year colleges. Therefore, he requested a "not" logic to eliminate the word "junior" from the abstracts. Many programs similar to the one proposed had been operated at junior colleges, but they were of no interest to the planners of the new project. The search resulted in the elimination of nine records which described programs at junior colleges.

Once again a decision was made to iterate. The prefix for "achievement"

Figure 10

Third Sequence of a Simple Search

```
A. title < urban>
        TITLE -
B. $ < negro>
C. c.
C. $ < urban>
C. abstr < urban>
        ABSTRACT -
D. $ < negro>
E.
LOGIC a + b + c + d
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED      145
 SELECTED       48      SUBSET 3
 VARIABLES SATISFIED
      A          7
      B          2
      C         22
      D         29

ITERATE?
yes
```

Figure 11

Fourth Sequence of a Simple Search

```
A. abstr < junior>
       ABSTRACT -
B.
LOGIC not a
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED        48
 SELECTED        39       SUBSET 4
 VARIABLES SATISFIED
     A            9
```

was entered into titles and abstracts (see Figure 12). The logic requested was an "or" logic; the word "achievement" could appear in either the title or the abstract. A second logic statement was entered using an "and" logic. This was labeled statement Logic A. Request for a third logic was rejected by the user. A search of subset four, which contained only 39 records, selected 10 which reported on student achievement. Of these, the second logic statement indicated that only two of them showed the term "achievement" in both the abstract and title. The decision was made at this point to print the ten remaining records and, therefore, the request for iteration received the reply "no." The user elected the command "print" and requested that the printing be done at the high-speed printer at the Merrick Center. He entered a label so that he could identify his records.

It should be noted that he could go back to an earlier subset if he wished to get additional records. However, at this point he felt that he had the essence of the material to answer the 11 questions.

A review of the questions indicates that he chose not to focus on personality characteristics but rather on academic achievement. A conversation with the requestors indicated that their first concern was that the students who would be selected for the program would survive in the academic community. Subsequently, some attention could be given to reactions other than academic achievement.

A Complex Search

The ability to return to previously established sets allows the system to engage in complex searches. It allows the user, for example, to group several questions and explore them sequentially after the

Figure 12

Final Sequence in a Simple Search

```
ITERATE?
yes
A. abstr < achievemen>
        ABSTRACT -
B. title <$
        TITLE -
C.
LOGIC a or b
    SEARCH
LOGIC a and b
    A
LOGIC

SEARCH BEGINNING


 SEARCHED       39
 SELECTED       10      SUBSET 5
 VARIABLES SATISFIED
     A           9
     B           3

 LOGIC SATISFIED
     A           2
ITERATE?
no

print
-------
TERMINAL OR PRINTER?
printer
ENTER A HEADING
ronald r. reeves   college of education      6-29-70
```

initial data base has been reduced. In the following example two questions were explored. Both were concerned with the college setting. The first was a broad inquiry. The user wished to know whether the concept of governance was used in connection with the position of a college dean. The second concerned the existence of special programs on the college level for black students, particularly those from a disadvantaged background. The third question concerned the efficiency of searching abstracts, titles and descriptors in contrast to searching only one part of the record.

Figure 13 shows the program for the original run. The term "college" was requested in the abstracts. The final "e" was dropped when the title and descriptors were searched. This would also permit identification of such words as "collegiate" or "colleges." The basic logic used the "or" pattern; a record would be identified if the words were found in any one of the three parts of the record. A second logic was entered in which a record would be selected only if the word "college" appeared in the abstract. The strategy was to exclude records where the term "college" or "collegiate" appeared in the title or descriptors. A comparison of the two logic statements would indicate the loss of efficiency that would occur if only the abstract had been searched. Finally, a third logic statement was entered in which it was required that the term appear in all three parts of the record. Additional logic statements could have been written to further specify the effectiveness of searching different parts. However, the three were deemed sufficient for the immediate purpose. It should be noted that a subset was constructed only for the first logic statement. In this case, the use of "or" logic should have provided the largest number of abstracts.

Figure 13

First Sequence of a Complex Search

```
select
-------
A. abstr < college>
       ABSTRACT -
B. title < colleg>
       TITLE -
C. descrip $
       DESCRIPTOR -
D.
LOGIC a or b or c
    SEARCH
LOGIC a and not (b or c)
    A
LOGIC a and b and c
    B
LOGIC

SEARCH BEGINNING


 SEARCHED   27026
 SELECTED    5830     SUBSET 1
 VARIABLES SATISFIED
     A        4957
     B        2000
     C        2995

 LOGIC SATISFIED
     A        2203
     B        1242
ITERATE?
yes
A. abstr < governan>
       ABSTRACT -
B. $ < dean>
C.
LOGIC a or b
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED    5830
 SELECTED     145     SUBSET 2
 VARIABLES SATISFIED
     A          55
     B          91

ITERATE?
y,1
```

The search which produced subset one identified 5,830 records which used the term "college" or the prefix "colleg" at some point in the record. An examination of the report indicated that the highest frequency of occurrence of the word was in the abstracts. However, there were 873 records in which the term appeared in either title or descriptor but did not appear in the abstract. This would indicate an increase of 14.97 percent in the number of records identified. If only the titles had been searched, only 2,000 records would have been located. This represents 34.31 percent of those identified on the search of all three elements. Similarly, if only the descriptors had been used, 2,995 records would have been located. This number is 51.37 percent of the total number identified. An index system, using only the descriptors, would have missed almost half of the records. Of course, there is no way of knowing without reading the abstracts whether the additional 49.63 percent would add significantly to the information which the user desired. The second and third logic statements shed additional light on the problem of the efficiency of searching only one part of the record. If the search had been made according to the second logic statement, in which the term was requested only in the abstract but in neither title nor descriptor, the number of records identified would have been reduced to 2,203. If the "and" logic had been employed as in the third statement so that the term would have to appear in all three elements of the record, the number would have been further reduced to 1,242.

The search was pursued concerning the first topic. The terms "governan" and "dean" were entered and a search requested of abstracts only. A total of 145 documents were identified. Fifty-five of them used some form of the term "governan"; 91 contained a reference to the dean.

Obviously, the two terms are frequently used in conjunction with one another. The request for "ITERATE" received an affirmative reply, but included a command to return to subset one (y,1) rather than to re-enter subset two to continue the search.

Since the second question was concerned with four-year colleges only, a first run removed references to junior colleges. This was achieved by using the "not" logic in the search. As can be seen from Figure 14, a total of 1,165 records were eliminated by this action. The next search required that both the term "test" and "achievement" be present in the abstract since the concern was with methods for objective evaluation. Subset three contained only 185 records. Note that there were many more references to "achievement" and "test" than there were records which contained both terms.

The next exploration (see Figure 15) entered the two terms "Negro" and "black." Although the terms have become synonymous in recent years, many of the ERIC records predate the wide usage of the term "black" as a synonym for "Negro." However, both terms were used in the exploration. The search reduced the data set to only seven records, six of which contained the term "Negro" and one of which contained the word "black." The decision was then made to examine the abstract containing the word "black." The single term was entered to select the one record. Figure 16 indicates the action that followed the search. When the request concerning "ITERATE?" was made, the user replied in the negative, using an "n" as an abbreviation for "no." He then requested that the one record be printed by entering the word "print." He also chose to have it printed at the remote access terminal so that he could examine it immediately. As he suspected, the

Figure 14

Second Sequence of a Complex Search

```
A. abstr < junior>
       ABSTRACT -
B.
LOGIC not a
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED    5830
 SELECTED    4665      SUBSET 2
 VARIABLES SATISFIED
     A       1165

ITERATE?
yes
A. abstr < achievemen>
       ABSTRACT -
B. $ < test>
C.
LOGIC a and b
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED    4665
 SELECTED     185      SUBSET 3
 VARIABLES SATISFIED
     A        404
     B        910

ITERATE?
yes
```

Figure 15

Third Sequence of a Complex Search

```
A. abstr < negro>
       ABSTRACT -
B. $ < black>
C.
LOGIC a or b
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED      185
 SELECTED        7      SUBSET 4
 VARIABLES SATISFIED
     A           6
     B           1

ITERATE?
yes
A. abstr < black>
       ABSTRACT -
B.
LOGIC a
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED        7
 SELECTED        1      SUBSET 5
 VARIABLES SATISFIED
     A           1
```

Figure 16

Printout at the Terminal

```
ITERATE?
n

print
-------
TERMINAL OR PRINTER?
terminal

 ACCESSION NUMBER - ED029503
 CLEARINGHOUSE ACCESSION NUMBER - EM007215
                    PUBLICATION DATE - JUN 68
                    INSTITUTION CODE - SYN71980
                    EDRS PRICE - EDRS PRICE MF-$0.25 HC-$0.75
                    DESCRIPTIVE NOTE - 13P.
                    ISSUE -  RIEOCT69
INSTITUTION NAME - PENNSYLVANIA STATE UNIV., UNIVERSITY PARK.
  DIV. OF INSTRUCTIONAL SERVICES.
TITLE -  AN EXPERIMENTAL STUDY OF THE USE OF VISUAL ILLUSTRATIONS
  USED TO COMPLEMENT ORAL INSTRUCTION ON TELEVISION.
AUTHOR - DWYER, FRANCIS M., JR.
ABSTRACT - FIVE SLIDE SEQUENCES, EACH CONTAINING 39 BLACK-AND-
  WHITE SLIDES DESIGNED TO COMPLEMENT ORAL INSTRUCTION, AND
  CARRYING A 32 MINUTE ORAL INSTRUCTIONAL UNIT ON THE HEART,
  WERE PRESENTED TO 269 COLLEGE STUDENTS IN FIVE GROUPS THROUGH
  A TELEVISION RECEIVER.  THE PURPOSE WAS TWOFOLD: TO DETERMINE
  IF REDUNDANT INFORMATION PRESENTED SIMULTANEOUSLY THROUGH EYE
  AND EAR RESULTS IN MORE EFFECTIVE LEARNING THAN IF THE
  INFORMATION IS PRESENTED THROUGH THE EAR ALONE, AND TO MEASURE
  THE RELATIVE EFFECTIVENESS OF VARIED TYPES OF VISUAL
  ILLUSTRATIONS USED TO FACILITATE STUDENT REALIZATION OF VARIED
  EDUCATIONAL OBJECTIVES. THE RESULTS OF THE TESTS INDICATED
  THAT VISUAL AIDS ARE EFFECTIVE WHEN LEARNING OBJECTIVES ARE
  SIMILAR TO THOSE MEASURED BY A DRAWING TEST, BUT THAT THEY
  ARE UNNECESSARY AND EVEN DISTRACTING WHEN THE LEARNING
  OBJECTIVES ARE SIMILAR TO THOSE MEASURED BY TERMINOLOGY,
  IDENTIFICATION, COMPREHENSION, AND TOTAL CRITERION TESTS.
  THIS MAY BE BECAUSE COLLEGE STUDENTS HAVE ALREADY BEEN
  SELECTED FOR THEIR VERBAL AND CONCEPTUAL ABILITY. IN
  TELEVISING PRESENTATIONS, THEREFORE, FOR COLLEGE LEVEL
  INSTRUCTION, VISUAL MATERIALS SHOULD BE USED ONLY INSOFAR AS
  THEY ARE JUSTIFIED BY LEARNING OBJECTIVES AND ACHIEVEMENT. (GO)
DESCRIPTOR - *AUDIOVISUAL INSTRUCTION; AURAL LEARNING; *HIGHER
  EDUCATION; *INSTRUCTIONAL TELEVISION; *MULTISENSORY LEARNING;
  PRODUCTION TECHNIQUES; SLIDES; *STUDENT CHARACTERISTICS;
  VERBAL LEARNING; VISUAL LEARNING
```

term "black" was not used as a synonym for "Negro"; rather it was concerned with black and white slides used in a teaching unit. After examining the abstract, the user initiated a request for iteration (see Figure 17) and was asked to specify the subset to which he wished to return. He asked to return to subset four in order to get the remaining abstracts. When he once again had the set of six identified, he requested the machine to list only four elements from the record: the accession number, the title, the author and the abstract; the printing of the other elements of the record was suppressed. Once again he requested a display at the terminal. This is shown in Figure 17. In order to prevent the production of large amounts of unwanted material and yet to permit the immediate evaluation, GIPSY is programmed to print 20 lines. If the user finds it helpful and desires more, he can get an additional 20 lines by simply returning the carriage on the remote access unit. He can, thus, explore as far as he wants. In this instance, after the first 20 lines, it was decided to terminate the exploration and to return to subset three. A subsequent question about the number of specialized programs concerning "test" and "achievement" in colleges in California was explored. The search indicated that eleven records met the criteria. The user at this point was satisfied and ended the search (see Figure 18).

Figure 17

Display at the Terminal

```
iterate
-------
SPECIFY SUBSET NO.
4
A. abstr <negro>
       ABSTRACT -
B.
LOGIC a
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED          7
 SELECTED          6      SUBSET 5
 VARIABLES SATISFIED
    A              6

ITERATE?
no

list
-------
TERMINAL OR PRINTER?
terminal
ENTER LABEL(S)
id
title
author
abstr
/

ACCESSION NUMBER - ED002218
TITLE -  CIVIC GROUP EXPERIMENTS IN COMPENSATORY DEVELOPMENT OF
  CULTURALLY DISADVANTAGED CHILDREN.
AUTHOR - BRAZZIEL, WILLIAM F.
ABSTRACT - A CIVIC GROUP'S EFFORTS TO PROVIDE OPPORTUNITIES FOR
  CULTURALLY DISADVANTAGED CHILDREN ARE DESCRIBED AND THE
  RESULTS OF THE PROGRAM ARE EVALUATED.  THE POOR ACADEMIC
  PERFORMANCE OF LOCAL NEGRO CHILDREN IN A STATE-WIDE TESTING
  PROGRAM STIMULATED ACTION AND RESULTED IN A GROUP COMPOSED OF
  PROFESSIONALS AND SEMIPROFESSIONALS FROM LOCAL, STATE, AND
  FEDERAL GOVERNMENT AGENCIES TEACHERS FROM THE LOCAL NEGRO
  COLLEGES PUBLIC SCHOOL TEACHERS AND PRINCIPALS MINISTERS
  CHURCH EDUCATION COMMITTEE CHAIRMAN AND HOUSEWIVES. THE GROUP
  IS SPONSORED BY NEGRO CIVIC ORGANIZATIONS AND A NEGRO
  NEWSPAPER. THE PROGRAM IS VARIED AND THE WORK IS CARRIED ON
  BY VOLUNTEERS. COMMITTEES SPONSOR SUCH VARIOUS ACTIVITIES AS
  REMEDIAL AND ENRICHMENT CLINICS, SUMMER SCHOOLS, INDIVIDUAL
  TUTORING, RECOGNITION OF MERITORIOUS SCHOOL ACHIEVEMENT,
  RESEARCH, EVALUATION, ADULT EDUCATION, ENCOURAGEMENT OF
  AESTHETIC ACTIVITY, AND AID TO GIFTED CHILDREN. FAVORABLE
```

Figure 18

Final Sequence in a Complex Search

```
iterate
-------
SPECIFY SUBSET NO.
3
A. abstr < california>
       ABSTRACT -
B.
LOGIC a
    SEARCH
LOGIC

SEARCH BEGINNING


 SEARCHED      185
 SELECTED       11      SUBSET 4
 VARIABLES SATISFIED
     A          11

ITERATE?
no

message
-------
I have run out of questions.  please send print outs.  thanks.
OK
end
-------
 G I P S Y - UNIVERSITY OF OKLAHOMA  MONDAY  JUNE 29, 1970
```

PART VI

Batch Processing GIPSY/ERIC

With the rapid increase in the number of records on the ERIC file and with the addition of the CIJE materials, searches of the total base will soon become unwieldy. One solution is to create smaller but comprehensive record files for searching through teleprocessing method. GIPSY contains the facility whereby smaller but carefully targeted record sets can be established. The sets are established through batch processing and may be stored on disc packs for immediate use or on tape for longer term storage.

In designing a special record set, the user writes a program quite similar to that which he would enter at the remote access keyboard. However, for batch processing, he must write the entire program which will be keypunched and submitted to the computing center in a fashion similar to that through which any other program would be handled.

Some minor changes are required in the GIPSY/ERIC routines for batch processing. The commands and logic statements are the same; however, the user must enter all commands ("form," "select," "iterate" etc.) beginning in Column 1; labels and variable entries begin in Column 2. The user can design his search and control the computer through the use of commands just as he would with the remote access unit. However, he does not get a response from the computer and, therefore, cannot alter his program while it is running.

In order to create a selected record file for future searches

through the remote access terminal, the user must contact the operator at the computing center and make arrangements for one or more disc packs to be available to receive the final "copy" or "dump" command. For later processing through the remote access or teleprocessing unit the command "dump" will be used. For processing by other means, it may be necessary to use the command "copy" in order to gain a fixed format field.

When the specific file has been created and the user wishes to search it from the remote terminal, he must notify the operator at the computing center which of the files he wishes to use. He may do so through the "message" function or in a direct telephone conversation.

The only advantage of batch processing is to reduce the number of records in the initial search. Thus, it is extremely important that the program for batch processing be written in such a fashion that all suitable records will be available when the interactive search from the remote access unit begins.

BIBLIOGRAPHY

Biography 3: George Boole, The sciences, 1965, 4, 8-11.

Blai, Boris, Jr. Why educational research? Educational media, 1970, 3, 2, 10-14.

Bruner, Jerome S. The process of education. Cambridge, Massachusetts: Harvard University Press, 1960.

Stephens, J. M. The process of schooling, New York: Holt, 1967.

APPENDIX

APPENDIX A

Current ERIC Clearinghouses and Their Scopes, July, 1970

ADULT EDUCATION Syracuse University 107 Roney Lane Syracuse, New York 13210	ERIC/AE is responsible for research and other documents on formal and informal adult and continuing education in all settings.
COUNSELING AND PERSONNEL SERVICES 611 Church Street Ann Arbor, Michigan 48104	ERIC/CAPS focuses on information relevant to personnel work at all levels and in all settings, including college student personnel work, school psychology, school social work, elementary and secondary school counseling, school health work, school psychiatry, employment counseling, and personnel work research. Included are materials on pupil, student, and adult characteristics; educational, occupational, and community settings; and the types of assistance provided by personnel workers.
THE DISADVANTAGED Teachers College--Box 40 Columbia University New York, New York 10027	ERIC/IRCD is responsible for research reports and other documents on the educational, psychological, social, and general development of urban children and youth who are socially or economically disadvantaged.
EARLY CHILDHOOD EDUCATION University of Illinois 805 West Pennsylvania Avenue Urbana, Illinois 61801	ERIC/ECE is responsible for research documents on the physiological, psychological, and cultural development of children from birth through the primary grades, with major focus on educational theory, research and practice related to the development of young children.
EDUCATIONAL ADMINISTRATION 320 Hendricks Hall University of Oregon Eugene, Oregon 97403	ERIC/CEA's subject area is the administration of educational organizations on the elementary and secondary levels, including educational facilities.
EDUCATIONAL MEDIA AND TECHNOLOGY Institute for Communication Research Stanford University Stanford, California 94305	ERIC at Stanford is responsible for information on application of new media and technological innovation to education, including such subjects as instructional television, computer-assisted instruction, and programmed learning.

EXCEPTIONAL CHILDREN The Council for Exceptional Children 1411 S. Jefferson Davis Highway Suite 900 Arlington, Virginia 22202	ERIC/CEC acquires selected documents concerning children and youth who require special services. Included are the visually, aurally, and mentally handicapped, the emotionally disturbed and physically handicapped, children with learning disabilities, and those with speech defects.
HIGHER EDUCATION George Washington University One Dupont Circle--Suite 630 Washington, D.C. 20036	ERIC/CHE is responsible for research documents on higher education, with the exception of reports on both teacher education and teaching English in higher education.
JUNIOR COLLEGES University of California at Los Angeles 405 Hilgard Avenue Los Angeles, California 90024	The Clearinghouse is responsible for research documents about public and private community and junior colleges, including studies on students, staff, curricula, programs, libraries, and community services.
LIBRARY AND INFORMATION SCIENCES American Society for Information Science 1140 Connecticut Avenue, NW Suite 804 Washington, D.C. 20036	ERIC/CLIS is responsible for research documents on the operation of libraries and information centers, the technology used to improve their operations, and the education and training of library and information specialists.
LINGUISTICS Center for Applied Linguistics 1717 Massachusetts Avenue,NW Washington, D.C. 20036	CAL/ERIC is responsible for research reports on linguistics and all related language sciences, uncommonly taught languages, the teaching of English as a foreign or second language, and the teaching of English as a native language to speakers of non-standard dialects.
READING 200 Pine Hall School of Education Indiana University Bloomington, Indiana 47401	ERIC/CRIER focuses on information related to all aspects of reading behavior with emphasis on physiology, psychology, sociology, and the teaching of reading. Included are reports on the development and evaluation of instructional materials, curricula, tests and measurements, preparation of reading teachers and specialists, and methodology at all levels; the role of libraries and other agencies in fostering and guiding reading; and diagnostic and remedial services in school and clinic settings.

RURAL EDUCATION AND SMALL SCHOOLS New Mexico State University Box 3 AP Las Cruces, New Mexico 88001	ERIC/CRESS is responsible for information on organization, administration, curriculum, instruction, innovative programs, and other aspects of small schools, rural education, outdoor education, migrant education, Indian education, and Mexican American education.
SCIENCE AND MATHEMATICS EDUCATION Ohio State University 1460 West Lane Avenue Columbus, Ohio 43221	ERIC/SMAC is responsible for reports on all levels of science and mathematics education, and on adult and continuing education in science and mathematics. Included also are materials and programs related to environmental education.
SOCIAL SCIENCE EDUCATION 970 Aurora Avenue Social Science Building Boulder, Colorado 80302	ERIC/CSSE is responsible for information on social studies, social science education, and social education.
TEACHER EDUCATION One Dupont Circle--Suite 616 Washington, D.C. 20036	The Clearinghouse focuses on information about preservice and in-service preparation of teachers, supporting school personnel, and teacher educators.
TEACHING OF ENGLISH National Council of Teachers of English 508 South Sixth Street Champaign, Illinois 61820	NCTE/ERIC focuses on research reports and other documents relative to all aspects of the teaching of English from kindergarten through college, and the preparation of the teachers of English and of specialists in English education.
TEACHING OF FOREIGN LANGUAGES Modern Language Association of America 62 Fifth Avenue New York, New York 10011	MLA/ERIC is responsible for research documents on teaching French, German, Italian, Russian, Spanish, Latin, and classical Greek at all instructional levels. It also is responsible for documents on the teaching of English in undergraduate and graduate education.
TESTS, MEASUREMENT, AND EVALUATION Educational Testing Service Rosedale Road Princeton, New Jersey 08540	ERIC/TM is responsible for documents presenting descriptions of: (1) tests and other measurement devices; (2) evaluation procedures and techniques; and (3) the application of tests, measurement, or evaluation in educational projects or pgorams.

VOCATIONAL AND TECHNICAL EDUCATION
Ohio State University
1900 Kenney Road
Columbus, Ohio 43212

The Clearinghouse focuses on research documents and related resources in vocational and technical education, new sub-professional fields, and the related fields of industrial arts education, manpower economics, occupational psychology, and occupational sociology.

Provided by the Clearinghouse on Rural Education and Small Schools--July, 1970

APPENDIX B

INFORMATION SYSTEMS AND EVALUATION CENTER
UNIVERSITY OF OKLAHOMA
Norman, Oklahoma

The University of Oklahoma's Merrick Computing Center and the College of Education have established the Information Systems and Evaluation Center (ISEC). The center provides a variety of services concerned with program evaluation, systems analysis and information sources.

Among its activities is a computerized literature search. Currently, RIE (Research in Education) and CIJE (Current Index to Journals in Education) records are available. These are provided by the 20 ERIC clearinghouses and are published monthly by the U. S. Office of Education. The RIE file makes available abstracts on conferences, research projects, demonstration programs and other significant educational activities; the CIJE file provides access to the mainstream of the professional literature.

The search service is available to schools, colleges and universities developing special projects or applications for federal funds, to graduate students approaching a thesis or dissertation and to other professionals with specific interests in educational research and development.

How to Obtain a Search

A search of the files may be obtained by writing to ISEC. A pre-search form is included for your convenience. The most efficient process is for you to write your question in standard English as precisely as you can.

In most instances, it is possible through the iterating nature of

GIPSY to select a few highly relevant documents. The average run produces about 27 abstracts. Not all of them will be significant. Because of the complexity of the English language, some abstracts will be selected which contain the correct words but with a different meaning or within another context. If a search of the ERIC abstracts shows that nothing on the topic has been reported, valuable information has been gained; the absence of information can be just as significant as its presence.

Abstracts will usually be mailed within a day or two after the request is received; the billing process requires about ten days. The cost of a search is based upon the amount of computer time required. At present, an average run costs about $45.

If you desire a search or additional information, please address your request to:

Information Systems and Evaluation Center
College of Education
820 Van Vleet Oval
Norman, Oklahoma 73069

INFORMATION SYSTEMS AND EVALUATION CENTER

THE UNIVERSITY OF OKLAHOMA

Merrick Computing Center

College of Education
820 Van Vleet Oval
Norman, OK 73069

GIPSY/ERIC
PRE SEARCH INFORMATION

Name_____Address_____

School_____City, State, Zip_____

Department_____Phone_____

Please write your question here. Use standard English. Be sure to include all significant constraints such as grade level or subject area.

_____

_____

_____

_____

1. Name and address of person to be billed:

Name_____Address_____

City, State, Zip_____

2. Purpose of search:

_____

FOR OFFICE USE ONLY

Date received _____ Number of Abstracts _____

Date searched _____ Time _____

Date mailed _____

ISEC 7020271o

UNIVERSITY OF OKLAHOMA

Information Science Series

No. 1
Linguistic Concordance System Manual

Joseph E. Grimes
James W. Sweeney
Robert W. Shields
Larry F. Core

No. 2
General Information Processing System
Permian Palynology of North America and
Some Associated Problems

L. R. Wilson
Jack L. Morrison
William E. Reid

No. 3
General Information Processing System
Remote Terminal Users Guide

Charles H. Addison
Philip W. Blackwell
Wayne E. Smith, Jr.
Robert W. Shields
James W. Sweeney

No. 4
General Information Processing System
Application Description

Charles H. Addison
M. Duane Coney
Margaret Jones
Robert W. Shields
James W. Sweeney

No. 5
General Information Processing System
The GIPSY/ERIC Abstract Retrieval System

Gerald T. Kowitz
Ronald R. Reeves
Jerry R. Prather